EFFICIENT DERIVATIVE PRICING BY THE EXTENDED METHOD OF MOMENTS

P., GAGLIARDINI[*], C., GOURIEROUX[†] and E., RENAULT[‡]

First version: July 2004
This version: October 2010

[*]University of Lugano and Swiss Finance Institute.
[†]CREST, CEPREMAP (Paris) and University of Toronto.
[‡]CIRANO-CIREQ (Montreal) and University of North Carolina at Chapel Hill.

Acknowledgements: We thank D. Andrews, X. Chen, F. Corsi, R. Davidson, J. Fan, R. Garcia, J. Jackwerth, J. Jasiak, O. Linton, L. Mancini, A. Melino, C. Ortelli, O. Scaillet, M. Stutzer, F. Trojani, A. Vedolin, B. Werker, E. Zivot, a co-editor and anonymous referees for helpful comments. The first author gratefully acknowledges financial support of the Swiss National Science Foundation through the NCCR FINRISK network. The second author gratefully acknowledges financial support of NSERC Canada and of the chair AXA/Risk Foundation: "Large Risks in Insurance".

Efficient Derivative Pricing by the Extended Method of Moments

Abstract

In this paper we introduce the Extended Method of Moments (XMM) estimator. This estimator accommodates a more general set of moment restrictions than the standard Generalized Method of Moments (GMM) estimator. More specifically, the XMM differs from the GMM in that it can handle not only uniform conditional moment restrictions (i.e. valid for any value of the conditioning variable), but also local conditional moment restrictions valid for a given fixed value of the conditioning variable. The local conditional moment restrictions are of special relevance in derivative pricing for reconstructing the pricing operator at a given day, by using the information in a few cross-sections of observed traded derivative prices and a time series of underlying asset returns. The estimated derivative prices are consistent for large time series dimension, but fixed number of cross-sectionally observed derivative prices. The asymptotic properties of the XMM estimator are nonstandard, since the combination of uniform and local conditional moment restrictions induces different rates of convergence (parametric and nonparametric) for the parameters.

Keywords: Derivative Pricing, Trading Activity, GMM, Information Theoretic Estimation, KLIC, Identification, Weak Instrument, Nonparametric Efficiency, Semiparametric Efficiency.

JEL number: C13, C14, G12.

1 Introduction

The Generalized Method of Moments (GMM) was introduced by Hansen (1982) and Hansen, Singleton (1982) to estimate a structural parameter θ identified by Euler conditions:

$$p_{i,t} = E_t \left[M_{t,t+1}(\theta) p_{i,t+1} \right], \;\; i = 1, ..., n, \; \forall t, \tag{1.1}$$

where $p_{i,t}$, $i = 1, ..., n$, are the observed prices of n financial assets, E_t denotes the expectation conditional on the available information at date t, and $M_{t,t+1}(\theta)$ is the stochastic discount factor. Model (1.1) is semi-parametric. The GMM estimates parameter θ regardless of the conditional distribution of the state variables. This conditional distribution however becomes relevant when the Euler conditions (1.1) are used for pricing derivative assets. Indeed, when the derivative payoff is written on $p_{i,t+1}$ and a reliable current price is not available on the market, the derivative pricing requires the joint estimation of parameter θ and the conditional distribution of the state variables.

The Extended Method of Moments (XMM) estimator extends the standard GMM to accommodate a more general set of moment restrictions. The standard GMM is based on uniform conditional moment restrictions such as (1.1), that are valid for any value of the conditioning variables. The XMM can handle not only the uniform moment restrictions, but also local moment restrictions that are valid for a given value of the conditioning variables only. This leads to a new field of application of moment-based methods to derivative pricing, where the local moment restrictions correspond to Euler conditions for cross-sectionally observed prices of actively traded derivatives. Such a framework accounts for the fact that the trading activity is much smaller on each derivative than on the underlying asset, and that the characteristics of actively traded derivatives are not stable over time. More precisely, we explain how the XMM can be used for reconstructing the pricing operator on a given day, by using the information in a few cross-sections of observed actively traded derivative prices and a time series of underlying asset returns. To illustrate the principle of XMM, let us consider the S&P 500 index and its derivatives. Suppose an investor at date t_0 is interested in estimating the price $c_{t_0}(h, k)$ of a call option with time-to-maturity h and moneyness strike k that is currently not actively traded on the market. She has data on a time series of T daily returns of the S&P 500 index, as well as on a small cross-section of current option prices $c_{t_0}(h_j, k_j)$, $j = 1, ..., n$, of n highly traded derivatives. The XMM approach provides the estimated prices $\hat{c}_{t_0}(h, k)$ for different values of moneyness strike k and time-to-maturity h, that interpolate the observed prices of highly traded derivatives and satisfy the hypothesis of absence of arbitrage opportunities. These estimated prices are consistent for a large number of dates T, but a fixed, even small, number of observed derivative prices n.

To highlight the specificity of XMM with respect to GMM, we present in Section 2 an application to the S&P 500 index and its derivatives. First we show that the trading activity on the index option market is rather weak and the daily number of reliable derivative prices is small. Then we explain why the time series observations on the underlying index induce uniform moment restrictions, whereas the observed cross-sectional derivative prices correspond to local moment restrictions. The XMM estimator minimizes the discrepancy of the historical transition density from a kernel density estimator, subject to both types of moment restrictions. The estimation criterion includes a Kullback-Leibler information criterion associated with the local moment restrictions to ensure the compatibility of the estimated derivative prices with the absence of arbitrage opportunities. The comparison of the XMM estimator and the standard calibration estimator for S&P 500 options data shows clearly that XMM outperforms the traditional approach.

The theoretical properties of the XMM estimator are presented in Section 3 in a general semiparametric framework. We discuss the parameter identification under both uniform and local moment restrictions and derive the efficiency bound. We prove that the XMM estimator is consistent for a fixed number n of cross-sectional observations associated with the local restrictions and a large number T ($T \to \infty$) of time series observations associated with uniform restrictions. Moreover, the XMM estimator is asymptotically normal and semi-parametrically efficient. Section 4 concludes. The set of regularity assumptions and the proofs of the theoretical results are gathered in Appendix A. Proofs of technical Lemmas are given in Appendix B in the Supplemental Material [Gagliardini, Gouriéroux, Renault (2010)].

2 The XMM applied to derivative pricing

2.1 The activity on the index option market

In order to maintain a minimum activity, the Chicago Board Options Exchange (CBOE) enhances the market of options on the S&P 500 index by periodically issuing new option contracts [Hull (2005), p. 187]. The admissible times-to-maturity at issuing are 1-month, 2-month, 3-month, 6-month, 9-month, 12-month, etc, up to a maximal time-to-maturity. For instance, new 12-month options are issued every three months, when the options from the previous issuing attain the time-to-maturity of 9-month. This induces a cycle in the times-to-maturity of quoted options [see e.g. Schwartz (1987), Figure 1, and Pan (2002), Figure 2]. For any admissible time-to-maturity, the options are issued for a limited number of strikes around the value of the underlying asset at the issuing date.

This strategy restricts the number of options quoted on the market. Among these, only a few options are traded on a daily basis. This phenomenon is illustrated in Section 2.6, where we select the call and put options with a daily traded volume of more than 4000 contracts in June 2005. Since each contract corresponds to 100 options, 4000 contracts are worth between 5 millions USD and 7 millions USD, on average. The daily number of the highly traded options varies between a minimum of 7 in June 10, 2005 and a maximum of 31 in June 16, 2005. The corresponding times-to-maturity and moneyness strikes also vary in time. For instance in June 10, 2005 the actively traded times-to-maturity are 5, 70 and 135 days, while in June 16, 2005 the actively traded times-to-maturity are 1, 21, 46, 66, 131, 263 and 393 days. In brief, the number of highly traded derivatives on a given day is rather small. They correspond generally to puts for moneyness strikes below 1, and to calls otherwise. Moreover, the number of options and the moneyness strikes and times-to-maturity vary from one day to another due to the issuing cycle and the trading activity.

2.2 Calibration based on current derivative prices

The underlying asset features a regular trading activity, and the associated returns can be observed and are expected to be stationary. In contrast, the trading prices of a given option on two consecutive dates are not always observed and reliable. Therefore the associated returns might not be computed. Even if these option returns were available, they would be nonstationary due to the issuing cycle discussed above. To circumvent the difficulty in modelling the trading activity and its potential nonstationary effect on prices, the standard methodology consists in calibrating daily the pricing operator [1]. More precisely, let us assume that at date t_0 the option prices $c_{t_0}(h_j, k_j)$, $j = 1, ..., n$, are observed, and that a parametric model for the risk-neutral dynamics of the relevant state variables implies a pricing formula $c_{t_0}(h, k; \theta)$ for the option prices at t_0. The number n of actively traded derivatives and their design h_j, k_j depend on date t_0, but the time index is omitted for expository purpose. The unknown parameter θ is usually estimated daily by minimizing the least-squares criterion [2] [e.g., Bakshi et al. (1997)]:

$$\hat{\theta}_{t_0} = \arg\min_{\theta} \sum_{j=1}^{n} \left[c_{t_0}(h_j, k_j) - c_{t_0}(h_j, k_j; \theta) \right]^2 . \tag{2.1}$$

This practice has the following drawbacks: First, the estimated parameters $\hat{\theta}_{t_0}$ are generally erratic over time. Second, the approximated prices $\hat{c}_{t_0}(h_j, k_j) = c_{t_0}(h_j, k_j; \hat{\theta}_{t_0})$ are not compatible with the absence of arbitrage opportunities, since the approximated option prices $\hat{c}_{t_0}(h_j, k_j)$ differ from the observed prices $c_{t_0}(h_j, k_j)$ for highly traded options [3]. Third, even if the estimates can be shown to be consistent when n is large and the options characteristics (h_j, k_j) are well distributed [see e.g. Aït-Sahalia and Lo (1998)], in practice n is small and the estimates are not very accurate.

2.3 Semi-parametric pricing model

In general, the underlying asset is much more actively traded than each of its derivatives. For instance, the daily traded volume of a portfolio mimicking the S&P 500 index, such as

[1] Or, in calibrating daily the Black-Scholes implied volatility surface.

[2] Equivalently, a penalized Least-Squares (LS) criterion [Jackwerth (2000)], a local LS criterion under shape constraints [Aït-Sahalia, Duarte (2003)], a LS criterion under convolution constraints [Bondarenko (2003)], or an entropy criterion [Buchen, Kelly (1996), Stutzer (1996), Jackwerth, Rubinstein (1996)] in a nonparametric setting.

[3] When the pricer provides a price different from the market price, the trader can have the misleading impression that she can profit from an arbitrage opportunity, which does not exist in reality.

the SPDR, is about one hundred millions USD. It is possible to improve the above calibration approach by considering jointly the time series of observations on the underlying asset and the cross-sectional data on its derivatives. For this purpose the specification cannot be reduced to the risk-neutral dynamics only, but has to define in a coherent way the historical and risk-neutral dynamics of the variables of interest. Let us consider a discrete time model. We denote by r_t the logarithmic return of the underlying asset between dates $t-1$ and t. We assume that the information available to the investors at date t is generated by the random vector X_t of state variables, including the return r_t as the first component.

Assumption on the state variables: *The process (X_t) in $\mathcal{X} \subset \mathbb{R}^d$ is strictly stationary and time homogeneous Markov under the historical probability with transition density $f_{X_t|X_{t-1}}$.*

We consider a semi-parametric pricing model defined by the historical dynamics and a stochastic discount factor (sdf) [Hansen, Richard (1987)]. The historical transition density of process (X_t) is left unconstrained. We adopt a parametric specification for the sdf $M_{t,t+1}(\theta) = m(X_{t+1};\theta)$, where $\theta \in \mathbb{R}^p$ is an unknown vector of risk premia parameters. This model implies a semi-parametric specification for the risk-neutral distribution. For instance, the relative price at time t of a European call with moneyness strike k and time-to-maturity h written on the underlying asset is given by:

$$c_t(h,k) = E\left[M_{t,t+h}(\theta)(\exp R_{t,h} - k)^+ \mid X_t\right],$$

where $M_{t,t+h}(\theta) = m(X_{t+1};\theta)\cdots m(X_{t+h};\theta)$ is the sdf between dates t and $t+h$, and $R_{t,h} = r_{t+1} + \cdots + r_{t+h}$ is the return of the underlying asset in this period. The option price depends on both the finite-dimensional sdf parameter θ and the functional historical parameter $f_{X_t|X_{t-1}}$.

We are interested in estimating the pricing operator at a given date t_0, that is, the mapping that associates any European call option payoff $\varphi_{t_0}(h,k) = (\exp R_{t_0,h} - k)^+$ with its price $c_{t_0}(h,k)$ at time t_0, for any time-to-maturity h and any moneyness strike k.

Observability assumption: *The data consists of a finite number n of derivative prices $c_{t_0}(h_j,k_j)$, $j = 1,...,n$, observed at date t_0, and T serial observations of the state variables X_t corresponding to the current and previous days $t = t_0 - T + 1, \ldots, t_0$.*

In particular, the state variables in the sdf are assumed observable by the econometrician. The no-arbitrage assumption implies the moment restrictions for the observed asset prices.

2.4 The moment restrictions

The moment restrictions are twofold. The constraints concerning the observed derivative prices at t_0 are given by:

$$c_{t_0}(h_j, k_j) = E\left[M_{t,t+h_j}(\theta)(\exp R_{t,h_j} - k_j)^+ | X_t = x_{t_0}\right], \quad j = 1, \dots, n. \tag{2.2}$$

The constraints concerning the riskfree asset and the underlying asset are:

$$\begin{cases} E[M_{t,t+1}(\theta)|\ X_t = x] = B(t, t+1), \quad \forall x \in \mathcal{X}, \\ E[M_{t,t+1}(\theta) \exp r_{t+1}|\ X_t = x] = 1, \qquad \forall x \in \mathcal{X}, \end{cases} \tag{2.3}$$

respectively, where $B(t, t+1)$ denotes the price at time t of the short-term riskfree bond. The conditional moment restrictions (2.2) are local, since they hold for a single value of the conditioning variable only, namely the value x_{t_0} of the state variable at time t_0. This is because we consider only observations of the derivative prices $c_{t_0}(h_j, k_j)$ at date t_0. Conversely, the prices of the underlying asset and the riskfree bond are observed for all trading days. Therefore the conditional moment restrictions (2.3) hold for all values of the state variables. They are called the uniform moment restrictions. The distinction between the uniform and local moment restrictions is a consequence of the differences between the trading activities of the underlying asset and its derivatives. Technically, it is the essential feature of the XMM that distinguishes this method from its predecessor GMM.

There are $n + 2$ local moment restrictions at date t_0 given by:

$$\begin{cases} E[M_{t,t+h_j}(\theta)(\exp R_{t,h_j} - k_j)^+ - c_{t_0}(h_j, k_j)|X_t = x_{t_0}] = 0, \;\; j = 1, \dots, n, \\ E\left[M_{t,t+1}(\theta) - B(t_0, t_0 + 1)|X_t = x_{t_0}\right] = 0, \\ E[M_{t,t+1}(\theta) \exp r_{t+1} - 1|X_t = x_{t_0}] = 0. \end{cases} \tag{2.4}$$

Let us denote the local moment restrictions (2.2) as:

$$E\left[\tilde{g}(Y; \theta)|X = x_0\right] = 0, \tag{2.5}$$

where $Y_t = (X_{t+1}, \cdots, X_{t+\bar{h}})'$ is the $\tilde{d}$-dimensional vector of relevant future values of the state variables, $\bar{h}$ is the largest time-to-maturity of interest, $x_0 \equiv x_{t_0}$, and the time index is suppressed. Similarly, the uniform moment restrictions (2.3) are written as:

$$E\left[g(Y;\theta)|X = x\right] = 0, \quad \forall x \in \mathcal{X}. \tag{2.6}$$

Then, the whole set of local moment restrictions (2.4) corresponds to:

$$E\left[g_2(Y;\theta)|X = x_0\right] = 0, \tag{2.7}$$

where $g_2 = (g', \tilde{g}')'$. Since we are interested in estimating the pricing operator at a given date t_0, the value x_0 is considered as a given constant.

We assume that the sdf parameter θ is identified from the two sets of conditional moment restrictions (2.5) and (2.6). We allow for the general case where some linear combinations η_2^*, say, of the components of θ are unidentifiable from the uniform moment restrictions (2.6) on the riskfree asset and the underlying asset, only. The identification of these linear combinations requires local moment restrictions (2.5) on the cross-sectional derivative prices at t_0. Intuitively, these linear combinations η_2^* correspond to risk-premia parameters associated with risk factors that are not spanned by the returns of the underlying asset and the riskfree asset. This point is further discussed in Sections 2.7 and 3.2.

2.5 The XMM estimator

The XMM estimator presented in this section is related to the recent literature on the information based GMM [e.g., Kitamura, Stutzer (1997), Imbens, Spady, Johnson (1998)]. It provides estimators of both the sdf parameter θ and the historical transition density $f(y|x)$ of Y_t given X_t. By using the parameterized sdf, the information based estimator of the historical transition density defines the estimated state price density for pricing derivatives.

The XMM approach involves a consistent nonparametric estimator of the historical transition density $f(y|x)$, such as the kernel density estimator:

$$\hat{f}(y|x) = \frac{1}{h_T^{\tilde{d}}} \sum_{t=1}^{T} \tilde{K}\left(\frac{y_t - y}{h_T}\right) K\left(\frac{x_t - x}{h_T}\right) / \sum_{t=1}^{T} K\left(\frac{x_t - x}{h_T}\right), \tag{2.8}$$

where K (resp. $\tilde{K}$) is the d-dimensional (resp. $\tilde{d}$-dimensional) kernel, h_T is the bandwidth and (x_t, y_t), $t = 1, ..., T$, are the historical sample data. [4] Next, this kernel density estimator is improved by selecting the conditional pdf that is the closest to $\widehat{f}\,(y|x)$, and satisfies the moment restrictions, as defined below.

Definition 1: The *XMM estimator* $\left(\hat{f}^*\,(\cdot|x_0)\,, \hat{f}^*\,(\cdot|x_1)\,, ..., \hat{f}^*\,(\cdot|x_T)\,, \widehat{\theta}\right)$ consists of the functions $f_0, f_1, ..., f_T$ defined on $\mathcal{Y} \subset \mathbb{R}^{\tilde{d}}$, and the parameter value θ, that minimize the objective function:

$$L_T = \frac{1}{T}\sum_{t=1}^{T} \int \frac{\left[\widehat{f}(y|x_t) - f_t(y)\right]^2}{\widehat{f}(y|x_t)} dy + h_T^d \int \log\left[\frac{f_0(y)}{\widehat{f}(y|x_0)}\right] f_0(y)dy, \tag{2.9}$$

subject to the constraints:

$$\int f_t(y)dy = 1, \;\; t = 1, ..., T, \qquad \int f_0(y)dy = 1,$$
$$\int g\,(y;\theta)\, f_t(y)dy = 0, \;\; t = 1, ..., T, \qquad \int g_2\,(y;\theta)\, f_0(y)dy = 0. \tag{2.10}$$

The objective function L_T has two components. The first component involves the chi-square distance between the density f_t and the kernel density estimator $\widehat{f}(.|x_t)$ at any sample point x_t, for $t = 1, ..., T$. The second component corresponds to the Kullback-Leibler Information Criterion (KLIC) between the density f_0 and the kernel estimator $\widehat{f}(.|x_0)$ at the given value x_0. In addition to the unit mass restrictions for the density functions, the constraints include the uniform moment restrictions (2.6) written for all sample points, and the whole set of local moment restrictions (2.7) at x_0. The combination of two types of discrepancy measures is motivated by computational and financial reasons. The chi-square criterion evaluated at the sample points allows for closed form solutions $f_1\,(\theta)$, ..., $f_T\,(\theta)$ for a given θ (see Appendix A.2.1). Therefore, the objective function can be easily concentrated with respect to functions f_1, ..., f_T, which reduces the dimension of the optimization problem. The KLIC criterion evaluated at x_0 ensures that the minimizer f_0 satisfies the positivity restriction [see e.g. Stutzer (1996) and Kitamura, Stutzer (1997)]. The positivity

[4] For expository purpose, the dates previous to t_0, at which data on (X, Y) are available, have been re-indexed as $t = 1, ..., T$.

of the associated state price density at t_0 guarantees the absence of arbitrage opportunities in the estimated derivative prices. The estimator of $\hat{\theta}$ minimizes the concentrated objective function:

$$\mathcal{L}_T^c(\theta) = \frac{1}{T}\sum_{t=1}^{T} \widehat{E}\left(g(\theta)|x_t\right)' \widehat{V}\left(g(\theta)|x_t\right)^{-1} \widehat{E}\left(g(\theta)|x_t\right) - h_T^d \log \widehat{E}\left(\exp\left(\lambda\left(\theta\right)' g_2(\theta)\right)|x_0\right), \tag{2.11}$$

where the Lagrange multiplier $\lambda(\theta) \in \mathbb{R}^{n+2}$ is such that:

$$\widehat{E}\left[g_2(\theta)\exp\left(\lambda\left(\theta\right)' g_2(\theta)\right)|x_0\right] = 0, \tag{2.12}$$

for all θ, and $\widehat{E}\left(g(\theta)|x_t\right)$ and $\widehat{V}\left(g(\theta)|x_t\right)$ denote the expectation and variance of $g(Y;\theta)$, respectively, w.r.t. the kernel estimator $\widehat{f}(y|x_t)$. The first part of the concentrated objective function (2.11) is reminiscent from the conditional version of the continuously updated GMM [Ai, Chen (2003), Antoine, Bonnal, Renault (2007)]. The estimates $\hat{\theta}$ and $\lambda(\hat{\theta})$ are computed by writing the constrained optimization (2.11)-(2.12) as a saddle-point problem [see e.g. Kitamura, Stutzer (1997)] and applying a standard Newton-Raphson algorithm. Then the estimator of $f(y|x_0)$ is given by:

$$\widehat{f}^*(y|x_0) = \frac{\exp\left(\lambda\left(\hat{\theta}\right)' g_2(y;\hat{\theta})\right)}{\widehat{E}\left[\exp\left(\lambda\left(\hat{\theta}\right)' g_2(\hat{\theta})\right)|x_0\right]}\widehat{f}(y|x_0),\ y \in \mathcal{Y}. \tag{2.13}$$

This conditional density is used to estimate the pricing operator at time t_0.

Definition 2: The *XMM estimator of the derivative price* $c_{t_0}(h,k)$ is:

$$\hat{c}_{t_0}(h,k) = \int M_{t_0,t_0+h}(\hat{\theta})\left(\exp R_{t_0,h} - k\right)^+ \hat{f}^*\left(y|x_0\right)dy, \tag{2.14}$$

for any time-to-maturity $h \leq \bar{h}$ and any moneyness strike k. The estimator of the pricing operator density at time t_0 up to time-to-maturity $\bar{h}$ is $M_{t_0,t_0+\bar{h}}(\hat{\theta})\hat{f}^*\left(y|x_0\right)$.

The constraints (2.10) imply that the estimator $\hat{c}_{t_0}(h,k)$ is equal to the observed option price $c_{t_0}(h_j,k_j)$ when $h = h_j$ and $k = k_j$, $j = 1,...,n$. Moreover, the no-arbitrage restrictions for the underlying asset and the riskfree asset are perfectly matched at all sample dates. The large sample properties of estimators $\hat{\theta}$ and $\hat{c}_{t_0}(h,k)$ in Definitions 1 and 2 are examined in Section 3. The asymptotic analysis for $T \to \infty$ corresponds to long histories of the state variables before t_0 and a fixed number n of observed derivative prices at t_0, and is conditional on $X_{t_0} = x_0$ and the option prices observed at t_0. The estimators $\hat{\theta}$ and $\hat{c}_{t_0}(h, k)$ are consistent and asymptotically normal. [5] The linear combinations of θ that are identifiable from the uniform moment restrictions (2.6) on the riskfree asset and the underlying asset only, are estimated at the standard parametric rate $\sqrt{T}$. Any other direction η_2^* in the parameter space (see Section 2.4) and the derivative prices as well are estimated at the rate $\sqrt{Th_T^d}$ corresponding to nonparametric estimation of conditional expectations given $X = x_0$. The estimators of derivative prices are asymptotically semi-parametrically efficient for the informational content of the no-arbitrage restrictions. [6] Finally, the XMM estimator in Definition 1 does not account for the restrictions induced by the time homogeneous first-order Markov assumption. While this could induce an efficiency loss, it has the advantage to provide computationally tractable estimators. [7] The difficulty of accounting for the time-homogenous first-order Markov assumption is not specific to XMM. It concerns also the standard GMM, when the researcher adopts the first-order Markov assumption for the state process. Such an assumption is used to ensure that the derivative prices are functions of a finite-dimensional state vector and to derive the optimal instruments.

2.6 Application to S&P 500 options

In this section we compare the XMM estimation and a traditional cross-sectional calibration approach using the data on the S&P 500 options in June 2005 with daily trading volume

[5] The sample end-point condition $X_{t_0} = x_0$ is irrelevant for the asymptotic properties of the kernel density estimator (2.8) when the data are weakly serially dependent (see the mixing condition in Assumption A.5).

[6] Asymptotically equivalent estimators are obtained by replacing the integral w.r.t. the kernel density estimator in (2.14) by the discrete sum involved in a kernel regression estimator. These latter estimators involve a smoothing w.r.t. X only, and are used in the application in Section 2.6.

[7] The first-order Markov assumption implies the conditional independence of the future value X_{t+1} and the past history $\underline{X_{t-1}}$ given the current value X_t. This can be written as $E[\Psi_1(X_{t+1})\Psi_2(\underline{X_{t-1}})|X_t] - E[\Psi_1(X_{t+1})|X_t]E[\Psi_2(\underline{X_{t-1}})|X_t] = 0$, for any integrable functions Ψ_1, Ψ_2. These restrictions are not conditional moment restrictions. The additional assumption of time-homogeneity of the Markov process might be used to develop an adaptive estimation method. This extension is out of the scope of our paper.

larger than 4000 contracts (see Section 2.1). The OptionMetrics database contains daily data, where the trades of an option with given time-to-maturity and strike are aggregated within the day. Even if the high frequency trading option prices are not available, this database provides the closing bid/ask quotes. For a given day, these two quotes are rather close to each other, and to trading prices, for options with trading volume larger than 4000 contracts. The average of the closing bid/ask quotes is retained in the empirical analysis as a proxy for the trading price at the end of the day.

i) Cross-sectional calibration

The cross-sectional calibration is based on a parametric stochastic volatility model for the risk-neutral distribution Q. We assume that under Q the S&P 500 return is such that:

$$r_t = r_{f,t} - \frac{1}{2}\sigma_t^2 + \sigma_t \varepsilon_t^*, \tag{2.15}$$

where $r_{f,t}$ is the riskfree rate between $t-1$ and t, process (ε_t^*) is a standard Gaussian white noise and σ_t^2 denotes the volatility. The short-term riskfree rate $r_{f,t}$ is assumed deterministic. The volatility (σ_t^2) is stochastic, independent of shocks (ε_t^*) on returns and follows an Autoregressive Gamma (ARG) process [Gouriéroux and Jasiak (2006), Darolles, Gouriéroux and Jasiak (2006)], which is a discrete time Cox, Ingersoll and Ross (CIR) process [Cox, Ingersoll and Ross (1985)]. Hence, the joint model for (r_t, σ_t^2) is a discrete time analogue of the Heston model [Heston (1993)]. The risk-neutral transition distribution of the stochastic volatility is noncentral gamma and is more conveniently defined through its conditional Laplace transform:

$$E^Q\left[\exp\left(-u\sigma_{t+1}^2\right) \mid \underline{\sigma_t^2}\right] = \exp\left[-a^*(u)\sigma_t^2 - b^*(u)\right],\ u \geq 0, \tag{2.16}$$

where $E^Q[.]$ denotes the expectation under Q, and functions a^*, b^* are defined by $a^*(u) = \rho^* u/(1+c^*u)$ and $b^*(u) = \delta^* \log(1+c^*u)$. Parameter ρ^*, $\rho^* > 0$, is the risk-neutral first-order autocorrelation of volatility process (σ_t^2); parameter δ^*, $\delta^* \geq 0$, describes its (conditional) risk-neutral over-/under-dispersion; parameter c^*, $c^* > 0$, is a scale parameter. The conditional Laplace transform (2.16) is exponential affine in the lagged observation and the ARG process is discrete-time affine. We refer to Gouriéroux and Jasiak (2006) for an in-depth discussion of the properties of the ARG process.

The relative option prices are function of the current value of volatility σ_t^2 and of the three parameters $\theta = (c^*, \delta^*, \rho^*)$, that is, $c_t(h,k) = c\left(h, k; \theta, \sigma_t^2\right)$, say. Function c is computed by Fourier Transform methods as in Carr and Madan (1999) by exploiting the affine property of the joint process of excess returns and stochastic volatility. [8] The volatility σ_t^2 is calibrated daily jointly with θ by minimizing the mean square errors as in (2.1) [9]. The calibrated parameter $\hat{\theta}_{t_0}$ and volatility $\hat{\sigma}_{t_0}$ for the first ten trading days of June 2005 are displayed in Table I.

[Insert Table I : Calibrated parameters (cross-sectional approach)
for the S&P 500 options in June, 2005]

We also report the Root Mean Squared Errors $RMSE_{t_0} = \left\{\frac{1}{n_{t_0}} \sum_{j=1}^{n_{t_0}} \left[c\left(h_j, k_j; \hat{\theta}_{t_0}, \hat{\sigma}_{t_0}^2\right) - c_{t_0}(h_j, k_j)\right]^2\right\}^{1/2}$ as goodness of fit measure. We observe that the calibrated parameters $\hat{\delta}_{t_0}^*$, $\hat{\rho}_{t_0}^*$ and $\hat{c}_{t_0}^*$, and the goodness of fit measure vary in time and are quite erratic. The variation of the goodness of fit is due to a large extent to the small and time-varying number of derivative prices used in the calibration.

ii) XMM estimation

In the XMM estimation we consider the bivariate vector of state variables:

$$X_t = (\tilde{r}_t, \sigma_t^2)', \tag{2.17}$$

where $\tilde{r}_t := r_t - r_{f,t}$ is the daily logarithmic return from the S&P 500 index in excess of the riskfree rate, and σ_t^2 is an observable volatility factor. More specifically, σ_t^2 is the one-scale realized volatility computed from 30-minute S&P 500 returns [e.g., Andersen et al. (2003)]. The parameterized sdf is exponential affine:

$$M_{t,t+1}(\theta) = e^{-r_{f,t+1}} \exp\left(-\theta_1 - \theta_2\sigma_{t+1}^2 - \theta_3\sigma_t^2 - \theta_4\tilde{r}_{t+1}\right), \tag{2.18}$$

[8]Equations (5)-(6) in Carr, Madan (1999) are used to compute the option price as a function of time-to-maturity h and discounted moneyness $B(t, t+h)k$ (see Appendix B in the Supplemental Material). The term structure of riskfree bond prices is assumed exogeneous and is estimated at date t_0 by cubic spline interpolation of market yields at available maturities. This motivates our preference for a specification with time-varying deterministic interest rate compared to one with constant interest rate.

[9]It would also be possible to replace σ_t^2 by the observed realized volatility and calibrate w.r.t. θ only. We have followed the standard approach which provides smaller pricing errors.

where $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$. [10] This specification of the sdf is justified by the fact that the corresponding semi-parametric model for the risk-neutral distribution nests the parametric model used for cross-sectional calibration in Section i) above. More precisely, the risk-neutral stochastic volatility model (2.15)-(2.16) can be derived from an historical stochastic volatility model of the same type and the exponential affine sdf (2.18) with appropriate restrictions on parameter θ (see Section 3.2).

For each trading day t_0 of June 2005, we estimate by XMM the sdf parameter $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$ and 5 option prices $c_{t_0}(h, k)$ at a constant time-to-maturity $h = 20$ days and moneyness strikes $k = .96, .98, 1, 1.02, 1.04$. The options are puts for $k < 1$, and calls for $k \geq 1$. The estimator is defined as in Section 2.5, using the current and previous $T = 1000$ daily historical observations on the state variables, and the derivative prices of the actively traded S&P 500 options at t_0. We use a product Gaussian kernel and select two bandwidths for the state variables according to the standard rule of thumb [Silverman (1986)]. The estimation results for the first ten trading days of June 2005 are displayed in Table II.

[Insert Table II : Estimated sdf parameters and option prices (XMM approach)
for the S&P 500 options in June, 2005]

A direct comparison of the estimated structural parameters given in Tables I and II is difficult, since the parameters in Table I concern the risk-neutral dynamics whereas those in Table II concern the sdf. For this reason, we focus in Sections iii) and iv) below on a direct comparison of estimated option prices. Nevertheless, the following two remarks are interesting. First, the XMM estimated parameters are much more stable over time than the calibrated parameters, since they use the same historical information on the underlying asset. Second, from the computational point of view, the calibration method is rather time consuming since it requires the inversion of the Fourier Transform for all option prices, at each evaluation of the criterion function and its partial derivatives w.r.t. the parameters, and at each step of the optimization algorithm. On the contrary, the full XMM estimation takes about one minute on a standard computer.

[10] The option prices at time-to-maturity h depend on the pattern of the deterministic short rate between t and $t + h$ by means of the bond price $B(t, t + h)$ only. The term structure of bond prices is estimated from market yields as in Section 2.6 i).

iii) Static comparison of estimated option prices

In Figure 1 we display the estimated relative prices of S&P 500 options on June, 1 and June 2, 2005, as a function of the discounted moneyness strike $\tilde{k} = B(t, t+h)k$ and time-to-maturity h.

[Insert Figure 1 : Estimated call and put price functions for the S&P 500 options in June, 1, and June, 2, 2005]

The results for June, 1 (resp. June, 2) are displayed in the upper (resp. lower) panels. In the right panels, the solid lines correspond to XMM estimates, and in the left panels to calibrated estimates. Circles correspond to the observed prices of highly traded options. On June, 1, there are four highly traded times-to-maturity, which are 12, 57, 77, and 209-day, respectively. For the longest time-to-maturity $h = 209$, there is only one actively traded call option (resp., one put for $h = 57$ and two calls for $h = 77$). All remaining highly traded options correspond to time-to-maturity 12. When both the put and call options with identical moneyness strike and time-to-maturity are actively traded, we select the put if $\tilde{k} < 1$, and the call, otherwise. This is compatible with the procedure suggested by Aït-Sahalia and Lo (1998). After applying this procedure, we end up with 11 highly traded options in June, 1, and 8 highly traded options for the 3 times-to-maturity in June, 2, 2005. As discussed in Section 2.2, the calibration approach may produce different values for observed and estimated option prices while, by definition, these prices coincide within the XMM approach. Both calibration and XMM can also be used for pricing puts and calls that do not correspond to highly traded times-to-maturity and highly traded moneyness strikes. To show this, Figure 1 includes (dashed lines) the times-to-maturity 120 for June, 1, 2005, and 119 for June, 2, 2005, which are not traded. The calibration method assumes a parametric risk-neutral model, whereas the XMM risk-neutral model is semiparametric. Any specification error in the fully parametric pricing model is detrimental for the calibration approach, while the XMM approach is more flexible. An advantage of XMM is that, by construction, the estimated derivative prices coincide with the market prices for highly traded options, while the calibrated prices differ from the observed prices, and these discrepancies can be large. This is not due to the specific parametric risk-neutral model that was used. It would also occur if a more complicated parametric model, or the nonparametric approach proposed in Aït-Sahalia and Lo (1998), was used. By construction, the specific risk-neutral stochastic volatility model that underlies the calibration method produces smooth symmetric option pricing functions w.r.t. the log-moneyness [see Gouriéroux, Jasiak (2001), Chapter 13.1.5], with similar types of curvature when the time-to-maturity increases. In contrast, the XMM produces option pricing functions with different skew according to the time-to-maturity, which means that the method captures the leverage effect and its term structure.

iv) Dynamic comparison of estimated option prices

Let us now consider the dynamics of the option pricing function. In Figure 2 we display the time series of Black-Scholes implied volatilities at a fixed time-to-maturity $h = 20$ and moneyness strikes $k = .96, 1, 1.04, 1.06$ for all trading days in June, 2005.

[Insert Figure 2: Time series of implied volatilities for S&P 500 options in June, 2005]

The XMM implied volatilities are indicated by circles and the calibrated implied volatilities are marked by squares. The XMM implied volatilities are more stable over time because of the use of the historical information on the underlying asset. Since most of the highly traded options have moneyness strikes close to at-the-money, the calibration approach is rather sensitive to infrequently observed option prices with extreme strikes.

The sample means of the two time series of implied volatilities in each panel of Figure 2 are different. In particular, for $k = .96$ and $k = 1$, the XMM implied volatilities are on average larger than the calibrated ones, and smaller for $k = 1.04$ and $k = 1.06$. The reason is that the XMM approach captures the smirk, i.e. the skewness, in the implied volatility curve, while the calibrated model can reproduce either a smile, or a flat pattern only.

2.7 Discussion and possible extensions

We have illustrated above how to implement the XMM estimator for derivative pricing in a two-factor model with exponential affine sdf. In practice, the econometrician has to select the set of observable underlying factors and the set of asset prices, including reliable derivative prices, which are used for defining the uniform and local moment restrictions.

There exist arguments for introducing more factors. (i) For instance, if the riskfree interest rate is assumed stochastic, additional factors can be the riskfree rate, its realized volatility, and the realized covolatility between the index return and the riskfree rate. This would lead to a 5-factor model. (ii) If the stochastic volatility follows a Markov process of order q larger than 1, the first-order Markov assumption is recovered by introducing both the current and lagged volatilities σ_t^2 and $\sigma_{t-1}^2, \cdots, \sigma_{t-q+1}^2$ in the vector X_t. Pragmatic arguments such as tractability and robustness suggest to limit the number of factors, e.g. to a number smaller or equal to 3.

The XMM approach can be easily extended to account for a finite number of cross-sections of reliable observed derivative prices, e.g. the last 10 trading days [11]. This is achieved by including in criterion (2.9) a local KLIC component for each cross-section. However, there is no a-priori reason to overlook any reliable prices data. Thus, we could propose to consider the prices of actively traded derivatives for all previous dates and not only for a finite number of dates. By using this additional information one may expect to improve the convergence rate of the sdf parameters that are not identifiable from the uniform restrictions (2.3), and possibly achieve the parametric convergence rate for more θ parameters, although not necessarily for all of them. For instance, when the sdf involves risk premium parameters for extreme risk, the associated estimators could admit a smaller rate, since the derivatives with large strike are very unfrequently traded. In any case, the estimated option prices themselves will still have a nonparametric rate of convergence since the pricing model is semi-parametric. When an infinite number of local conditional moment restrictions corresponding to the sparse characteristics of traded derivatives are considered, we might aggregate the local moment restrictions to get the uniform restriction $\sum_{h\in\mathcal{H}}\sum_{k\in\mathcal{K}} I_t(h,k)E\left[M_{t,t+h}(\theta)\left(\exp R_{t,h}-k\right)^+ - c_t(h,k)|X_t\right] = 0$, where $I_t(h,k)$ is the activity indicator that is equal to 1 if option (h,k) is actively traded at date t, and equal to 0, otherwise, and the sums are over some sets $\mathcal{H}$ of times-to-maturity and $\mathcal{K}$ of moneyness strikes. The analysis of the statistical properties of the associated estimators will heavily depend on the choice of sets $\mathcal{H}$ and $\mathcal{K}$, and on the assumptions concerning the trading activity. A model for trading activity would have to account for the effect of periodic issu-

[11] Or the 10 past trading days with the values of the state variable closest to x_0.

ing of options by the CBOE, the observed activity clustering, the activity due to dynamic arbitrage strategies applied by some investors and the activity when some investors use the options as standard insurance products. The introduction of a realistic trading activity model would lead to a joint estimation of pricing and activity parameters, provide estimated option prices, and allow for the prediction of future activity. Accounting for the option data at all dates is an important extension of our work which is left for future research.

Let us finally discuss the central question of identification. We have pointed out in Section 2.4 that some sdf parameters may be unidentifiable, when the uniform moment restrictions (2.3) for the underlying asset and the riskfree asset only are considered. Intuitively, a lack of identification has to be expected, whenever the returns of the underlying asset and the riskfree asset do not span the set of state variables, and there are at least as many unknown sdf parameters as state variables. In such a framework, there may exist some directions of change in parameter θ that produce a change in the sdf $M_{t,t+1}(\theta)$ orthogonal to the underlying asset and riskfree asset gross returns. In Section 3.2 we formalize this intuitive argument for the specific stochastic volatility Data Generating Process (DGP) used in the application and prove lack of identification from uniform moment conditions.

It could be proposed to introduce parameter restrictions to solve the identification problem. For instance, the restriction $\theta_3 = 0$ in the sdf (2.18) yields identification of the full sdf parameter vector from the uniform moment conditions (2.3). From the analysis in Section 3.2 such a restriction is equivalent to imposing an a-priori level of the risk premium parameter θ_2 for stochastic volatility. In this case, the local moment restrictions from the derivative prices are noninformative for the estimation of θ, but still contribute to efficient estimation of the derivative prices. While imposing identifying restrictions simplifies the estimation problem, such an approach has some limitations. First, the degree and the directions of underidentification are not known a-priori since they depend on the unknown DGP. Second, an ad-hoc parametric restriction is likely misspecified and can lead to mispricing. Under the maintained hypothesis that the researcher knows the correct degree of underidentification, the validity of the selected identifying restrictions can be tested by a standard specification test for conditional moment restrictions. The difficulties in selecting the correct identifying restrictions are overcome by the XMM estimator, which is robust to the lack of identification from the uniform moment restrictions.

3 Theoretical properties of XMM

Let us now examine the theoretical properties of the XMM estimator introduced in Section 2 for derivative pricing. First we discuss the parameter of interest and the moment restrictions. Next we derive the identification conditions and explain how they differ from the standard GMM conditions. Then, we introduce instruments, define a convenient notion of efficiency, called kernel nonparametric efficiency, and derive the optimal instruments. Finally, we prove the kernel nonparametric efficiency of the XMM estimator.

3.1 The parameter of interest

Let us consider a semi-parametric estimation of the conditional moments:

$$E_0\left[a(Y;\theta_0)|X=x_0\right], \tag{3.1}$$

subject to the uniform and local conditional moment restrictions:

$$\begin{aligned} E_0\left[g(Y;\theta_0)|X=x\right] &= 0, \;\; \forall x \in \mathcal{X}, \;\; P_0\text{-a.s.}, & (3.2)\\ E_0\left[\tilde{g}(Y;\theta_0)|X=x_0\right] &= 0, & (3.3) \end{aligned}$$

where θ_0 is the true parameter value and $E_0[.]$ denotes the expectation under the true DGP P_0. The unknown parameter value θ_0 is in set $\Theta \subset \mathbb{R}^p$, variables (X, Y) are in $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^d \times \mathbb{R}^{\tilde{d}}$, and x_0 is a given value in $\mathcal{X}$. The moment functions g and $\tilde{g}$ are given vector-valued functions on $\mathcal{Y} \times \Theta$. Function a on $\mathcal{Y} \times \Theta$ has dimension L. In the derivative pricing application in Section 2, vector $a(Y;\theta)$ is the product of sdf $M_{t,t+h}(\theta)$ and payoff $(\exp R_{t,h} - k)^+$ on the L derivatives of interest. Vector functions g and $\tilde{g}$ define the uniform moment restrictions on the riskfree asset and underlying asset, and the local moment restrictions on observed derivative prices [see (2.5) and (2.6)].

The conditional moment of interest (3.1) can be viewed as an additional parameter β_0 in $B \subset \mathbb{R}^L$ identified by the local conditional moment restrictions:

$$E_0\left[a(Y;\theta_0) - \beta_0|X=x_0\right] = 0. \tag{3.4}$$

Thus, all parameters to be estimated are in the extended vector $\theta_0^* = \left(\theta_0', \beta_0'\right)'$ that satisfies the extended set of uniform and local moment restrictions (3.2), (3.3), (3.4) [see Back, Brown (1992) for a similar extension of the parameter vector]. In this interpretation, β_0 is the parameter of interest while θ_0 is a nuisance parameter.

3.2 Identification

Let us now consider the semi-parametric identification of extended parameter θ_0^*. From moment restriction (3.4), it follows that β_0 is identified if θ_0 is identified. Thus, we can restrict the analysis to the identification of θ_0.

i) The identification assumption

Assumption a.1: *The true value of parameter θ_0 is globally semi-parametrically identified:*

$$\left\{ \begin{array}{l} E_0\left[g(Y;\theta)|X=x\right]=0,\ \forall x\in\mathcal{X},\ P_0\text{-a.s.} \\ E_0\left[\tilde{g}(Y;\theta)|X=x_0\right]=0 \end{array} \right. ,\ \theta\in\Theta\ \Rightarrow\ \theta=\theta_0.$$

Assumption a.2: *The true value of parameter θ_0 is locally semi-parametrically identified:*

$$\left\{ \begin{array}{l} E_0\left[\dfrac{\partial g}{\partial \theta'}(Y;\theta_0)|X=x\right]\alpha=0,\ \forall x\in\mathcal{X},\ P_0\text{-a.s.} \\ E_0\left[\dfrac{\partial \tilde{g}}{\partial \theta'}(Y;\theta_0)|X=x_0\right]\alpha=0 \end{array} \right. ,\ \alpha\in\mathbb{R}^p\ \Rightarrow \alpha=0.$$

We need to distinguish the linear transformations of θ_0, $\alpha'\theta_0$, say, where $\alpha\in\mathbb{R}^p$, that are identifiable from the uniform moment restrictions (3.2) alone, from the linear transformations of θ_0 that are identifiable only from both uniform and local moment restrictions (3.2) and (3.3). The former transformations are called full-information identifiable, while the latter ones are called full-information unidentifiable. Let us consider the linear space:

$$\mathcal{J}^* = \left\{\alpha\in\mathbb{R}^p:\ E_0\left[\frac{\partial g}{\partial \theta'}(Y;\theta_0)|X=x\right]\alpha=0,\ \forall x\in\mathcal{X},\ P_0\text{-a.s.}\right\},$$

of dimension s^*, say, $0\le s^*\le p$. The full-information identified transformations are $\alpha'\theta_0$ with $\alpha\in(\mathcal{J}^*)^\perp$, while the full-information unidentifiable transformations are $\alpha'\theta_0$ with $\alpha\in\mathbb{R}^p\setminus(\mathcal{J}^*)^\perp$. There exist parameterizations of the moment functions such that $p-s^*$

components of the parameter vector are full-information identifiable, and s^* components are full-information unidentifiable. Indeed, let us consider a linear change of parameter:

$$\eta^* = \begin{pmatrix} \eta_1^* \\ \eta_2^* \end{pmatrix} = \begin{pmatrix} R_1'\theta \\ R_2'\theta \end{pmatrix}, \tag{3.5}$$

where $R^* = [R_1, R_2]$ is an orthogonal (p,p) matrix, R_1 is a $(p, p-s^*)$ matrix whose columns span $(\mathcal{J}^*)^\perp$, and R_2 is a (p, s^*) matrix whose columns span $\mathcal{J}^*$. The matrices R_1 and R_2 depend on the DGP P_0. The subvectors $\eta_1^* \in \mathbb{R}^{p-s^*}$ and $\eta_2^* \in \mathbb{R}^{s^*}$ are full-information identifiable, and full-information underidentified, respectively. Assumption a.2 is equivalent to matrix $E_0\left[\frac{\partial \tilde{g}}{\partial \theta'}(Y;\theta_0)|X=x_0\right] R_2$ having full column rank. Thus, a necessary order condition for local identification is that the number of local moment restrictions is larger than or equal to s^*, i.e. the dimension of the linear space $\mathcal{J}^*$ characterizing the full-information unidentifiable parameters.

The standard GMM considers uniform moment restrictions only, and assumes full-information identification for the full vector θ_0 under the DGP P_0. To illustrate the difference with our setting, let us assume that the DGP P_0 is in the parametric stochastic volatility model which is compatible with the parametric risk-neutral specification of Section 2.6 i) and the semi-parametric model of Section 2.6 ii). Specifically, the true historical distribution of $X_t = (\tilde{r}_t, \sigma_t^2)'$ is such that:

$$\tilde{r}_t = \gamma_0 \sigma_t^2 + \sigma_t \varepsilon_t, \tag{3.6}$$

where $\tilde{r}_t = r_t - r_{f,t}$ is the underlying asset return in excess of the deterministic riskfree rate, the shocks (ε_t) are $IIN(0,1)$ and the volatility (σ_t^2) follows an ARG process independent of (ε_t) with parameters $\rho_0 \in [0,1)$, $\delta_0 > 0$ and $c_0 > 0$. The transition of (σ_t^2) is characterized by the conditional Laplace transform $E_0\left[\exp(-u\sigma_{t+1}^2)|\underline{\sigma_t^2}\right] = \exp\left[-a_0(u)\sigma_t^2 - b_0(u)\right]$, $u \geq 0$, where $a_0(u) = \rho_0 u/(1+c_0 u)$ and $b_0(u) = \delta_0 \log(1+c_0 u)$. The sdf is given by:

$$M_{t,t+1}(\theta) = e^{-r_{f,t+1}} \exp\left(-\theta_1 - \theta_2 \sigma_{t+1}^2 - \theta_3 \sigma_t^2 - \theta_4 \tilde{r}_{t+1}\right), \tag{3.7}$$

with true parameter value $\theta_0 = (\theta_1^0, \theta_2^0, \theta_3^0, \theta_4^0)'$. Then, the true risk-neutral distribution Q is:

$$\tilde{r}_t = -\frac{1}{2}\sigma_t^2 + \sigma_t \varepsilon_t^*, \tag{3.8}$$

where $\varepsilon_t^* \sim IIN(0,1)$ and (σ_t^2) follows an ARG process independent of (ε_t^*) with parameters ρ_0^*, δ_0^* and c_0^*, that are functions of ρ_0, δ_0, c_0 and θ_0 given in (A.24). The no-arbitrage restrictions (2.3) written under the DGP P_0 are satisfied for the true value of the sdf parameter θ_0 if, and only if (see Appendix A.3.1):

$$\theta_1^0 = -b_0\left(\lambda_2^0\right), \quad \theta_3^0 = -a_0\left(\lambda_2^0\right), \quad \theta_4^0 = \gamma_0 + 1/2, \tag{3.9}$$

where $\lambda_2^0 = \theta_2^0 + \gamma_0^2/2 - 1/8$. The restrictions on θ_1^0 and θ_3^0 fix the predetermined component of the sdf, while the last equality in (3.9) relates θ_4^0 with the volatility-in-mean parameter γ_0. Parameter θ_2^0 is unrestricted.

The econometrician adopts a semiparametric framework with parametric sdf (3.7) and uses the uniform and local moment restrictions (2.2) and (2.3) to identify θ_0. From equations (3.9), parameter θ_4^0 is full-information identified, while parameters θ_1^0, θ_2^0 and θ_3^0 are full-information unidentifiable. Thus, the uniform moment restrictions (2.3) on the riskfree asset and the underlying asset are insufficient to identify the full parameter vector θ_0. The linear space defining the full-information identifiable transformations is characterized next.

Proposition 1. *When the DGP P_0 is compatible with (3.6)-(3.8):*

(i) The full-information identifiable transformations are $\theta_0'\alpha$ with $\alpha \in (\mathcal{J}^*)^\perp$, where:

$$\mathcal{J}^* = \left\{ \alpha \in \mathbb{R}^4 : E_0^Q \left[\left(\begin{array}{c} \exp r_{f,t+1} \\ \exp r_{t+1} \end{array} \right) \xi_{t+1}'\alpha | X_t = x \right] = 0 \,, \forall x \in \mathcal{X} \right\},$$

and $\xi_{t+1} = (1, \sigma_{t+1}^2, \sigma_t^2, \tilde{r}_{t+1})'$. The linear space $\mathcal{J}^*$ has dimension $s^* = 1$ and is spanned by vector $r_2 = \left(-\frac{db_0}{du}(\lambda_2^0), 1, -\frac{da_0}{du}(\lambda_2^0), 0\right)' = \left(-\frac{\delta_0 c_0}{1 + c_0\lambda_2^0}, 1, -\frac{\rho_0}{(1 + c_0\lambda_2^0)^2}, 0\right)'$.

(ii) The elements of the $(n, 1)$ vector $E_0\left[\partial\tilde{g}(Y;\theta_0)/\partial\theta' | X = x_0\right] r_2$ are given by:

$$(1-\rho_0^*) Cov_0^Q \left(\sigma_{t,t+h_j}^2, BS(k_j, \sigma_{t,t+h_j}^2) | X_t = x_0\right) + \rho_0^* Cov_0^Q \left(\sigma_{t+h_j}^2, BS(k_j, \sigma_{t,t+h_j}^2) | X_t = x_0\right),$$

for $j = 1, \cdots, n$, where $\sigma_{t,t+h}^2 = \sigma_{t+1}^2 + \cdots + \sigma_{t+h}^2$ is the integrated volatility between t and $t+h$, and $BS(k, \sigma^2)$ is the Black-Scholes price of a call option with time-to-maturity 1, moneyness strike k and volatility σ^2. Hence, the true value of the sdf parameter θ_0 is locally semi-parametrically identified by the conditional moment restrictions (2.2) and (2.3).

The characterization of linear space $\mathcal{J}^*$ in Proposition 1 (i) formalizes the intuitive spanning argument given in Section 2.7 to explain the lack of identification. Since $\xi_{t+1}'r_2 = \sigma_{t+1}^2 - E_0^Q[\sigma_{t+1}^2 | \sigma_t^2]$ (see A.27), the linear combination of state variables that is unspanned by the returns of the underlying asset and the riskfree asset, is the risk-neutral unexpected volatility. The dimension of the subspace of vectors $\alpha \in \mathbb{R}^4$ associated with the full-information identifiable transformations $\alpha'\theta_0$ is 3, while a full-information unidentifiable transformation is $\eta_2^* = R_2'\theta_0$, where $R_2 = r_2/\|r_2\|$. Since the Black-Scholes price is an increasing function of volatility, and the integrated volatility is stochastically increasing w.r.t. the spot volatility under Q conditionally on $X_t = x_0$ (see Lemma A.4 in Appendix A.3), from Proposition 1 (ii) the vector $E_0\left[\frac{\partial \tilde{g}}{\partial \theta'}(Y;\theta_0)|X = x_0\right] R_2$ has strictly positive elements. Thus, the full vector of parameters θ becomes identifiable when the local moment restrictions (2.2) from the observed derivative prices are taken into account.

ii) Admissible instrumental variables

It is also useful to discuss a weaker notion of identification, based on a given matrix of instruments $Z = H(X)$. The uniform conditional moment restrictions (3.2) imply the unconditional moment restrictions $E_0\left[Z \cdot g(Y;\theta_0)\right] =: E_0\left[g_1(X,Y;\theta_0)\right] = 0$. Let us denote the whole set of local conditional moment restrictions at x_0 as $E_0\left[g_2(Y;\theta_0)|X = x_0\right] = 0$, where $g_2 = (g', \tilde{g}')'$. Thus, parameter $\theta_0^* = \left(\theta_0', \beta_0'\right)'$ satisfies the moment restrictions:

$$E_0\left[g_1(X,Y;\theta_0)\right] = 0, \qquad (3.10)$$

$$E_0\left[g_2\left(Y;\theta_0\right) \mid X = x_0\right] = 0, \qquad (3.11)$$

$$E_0\left[a\left(Y;\theta_0\right) - \beta_0 \mid X = x_0\right] = 0. \qquad (3.12)$$

Definition 3: *The instrument Z is admissible* if the true value of the parameter θ_0 is globally semi-parametrically identified by the moment restrictions (3.10)-(3.11) and locally semi-parametrically identified by their differential counterparts.

Let us introduce the linear change of parameter:

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} R_{1Z}'\theta \\ R_{2Z}'\theta \end{pmatrix}, \qquad (3.13)$$

where $R = [R_{1,Z}, R_{2Z}]$ is an orthogonal (p,p) matrix, and R_{2Z} is a (p, s_Z) matrix whose columns span the null space $\mathcal{J}_Z =$ Ker $E_0\left[\frac{\partial g_1}{\partial \theta'}(X,Y;\theta_0)\right]$. The subvector of parameters η_1 is locally identified by the unconditional moment restrictions (3.10), whereas the subvector of parameters η_2 is identifiable only from both sets of restrictions (3.10) and (3.11).

3.3 Kernel moment estimators

Let Z be a given admissible instrument. Let us introduce a GMM-type estimator of θ_0^* that is obtained by minimizing a quadratic form of sample counterparts of moments (3.10)-(3.12). The kernel density estimator $\hat{f}(y|x)$ in (2.8) is used to estimate the conditional moments in (3.11) and (3.12):

$$\widehat{E}\left[g_2(Y;\theta)|x_0\right] := \int g_2(y;\theta)\hat{f}(y|x_0)dy \simeq \sum_{t=1}^{T} g_2(y_t;\theta)K\left(\frac{x_t-x_0}{h_T}\right) / \sum_{t=1}^{T} K\left(\frac{x_t-x_0}{h_T}\right),$$

and similarly for $E_0\left[a(Y;\theta)-\beta|X=x_0\right]$.

Definition 4: A *kernel moment estimator* $\widehat{\theta}_T^* = \left(\widehat{\theta}_T', \widehat{\beta}_T'\right)'$ of parameter $\theta_0^* = (\theta_0', \beta_0')'$ based on instrument Z is defined by:

$$\widehat{\theta}_T^* = \underset{\theta^*=\left(\theta',\beta'\right)'\in\Theta\times B}{\arg\min} \quad Q_T(\theta^*), \qquad Q_T(\theta^*) = \widehat{g}_T\left(\theta^*\right)' \, \Omega \, \widehat{g}_T\left(\theta^*\right),$$

where

$$\widehat{g}_T\left(\theta^*\right) = \left(\sqrt{T}\widehat{E}\left[g_1(X,Y;\theta)\right]', \sqrt{Th_T^d}\widehat{E}\left[g_2(Y;\theta)|x_0\right]', \sqrt{Th_T^d}\widehat{E}\left[a\left(Y;\theta\right)-\beta|x_0\right]'\right)',$$

$\widehat{E}[.]$ and $\widehat{E}\left[.|x_0\right]$ denote an historical sample average and a kernel estimator of the conditional moment, respectively, and Ω is a positive definite weighting matrix.

The empirical moments in $\widehat{g}_T(\theta^*)$ have different rates of convergence, that are parametric and nonparametric. This explains why the asymptotic analysis is different from the standard GMM. To derive the large sample properties of the kernel moment estimator $\widehat{\theta}_T^*$, we prove the weak convergence of a suitable empirical process derived from sample moments $\widehat{g}_T(\theta^*)$ (see Lemma A.1 in Appendix A.1.2). The definition of the empirical process is in the spirit of Stock, Wright (2000), but the choice of weak instruments is different. [12] The various rates of convergence in $\widehat{g}_T(\theta^*)$ can be disentangled by using the transformed parameters $(\eta_1', \eta_2')'$ defined in (3.13).

[12] The local moment restrictions (3.11) can be approximately written as $E_0\left[g_2\left(Y;\theta_0\right) \mid X=x_0\right] \simeq E_0\left[Z_T g_2\left(Y;\theta_0\right)\right] = 0$, where $Z_T = K\left(\frac{X-x_0}{h_T}\right) / \left[h_T^d f_X(x_0)\right]$ and f_X denotes the unconditional pdf of X. The "instrument" Z_T is weak in the sense of Stock, Wright (2000). However, it depends on T, and the rates of convergence of the estimators differ from the rates of convergence in Stock, Wright (2000).

Proposition 2. *Under Assumptions A.1-A.24 in Appendix A.1, the kernel moment estimator $\hat{\theta}_T^*$ based on instrument Z is consistent and asymptotically normal:*

$$\begin{pmatrix} \sqrt{T}\left(\widehat{\eta}_{1,T}-\eta_{1,0}\right) \\ \sqrt{Th_T^d}\left(\widehat{\eta}_{2,T}-\eta_{2,0}\right) \\ \sqrt{Th_T^d}\left(\widehat{\beta}_T-\beta_0\right) \end{pmatrix} \xrightarrow{d} N\left(B_\infty, \left(J_0'\Omega J_0\right)^{-1} J_0'\Omega V_0 \Omega J_0 \left(J_0'\Omega J_0\right)^{-1}\right),$$

as $T \to \infty$, where $(\eta_{1,0}', \eta_{2,0}')'$ is the true value of the transformed structural parameter, the bias is $B_\infty = -\sqrt{\bar{c}}\left(J_0'\Omega J_0\right)^{-1} J_0'\Omega b_0$, with $\bar{c} := \lim Th_T^{d+2m} \in [0,\infty)$ and $m \geq 2$ the order of differentiability of the pdf f_X of X, and matrices J_0, V_0, and vector b_0 are given in (A.2) and (A.4) in Appendix A.1, and depend on kernel K and instrument Z.

3.4 Kernel nonparametric efficiency bound

The class of kernel moment estimators helps us define a convenient notion of efficiency for estimating $\beta_0 = E_0(a|x_0) := E_0\left[a(Y;\theta_0)|X = x_0\right]$. Let us consider a scalar parameter β_0 and derive the optimal weighting matrix, admissible instruments and bandwidth. We assume that kernel K is a product kernel $K(u) = \prod_{i=1}^{d} \kappa(u_i)$ of order m.

i) Efficiency bound with optimal rate of convergence

Let us consider a bandwidth sequence $h_T = cT^{-\frac{1}{2m+d}}$, where $c > 0$ is a constant. From Proposition 2, it follows that estimator $\hat{\beta}_T$ achieves the optimal d-dimensional nonparametric rate of convergence $T^{-\frac{m}{2m+d}}$, and its asymptotic Mean Square Error (MSE) constant $M(\Omega, Z, c, a) > 0$ depends on the weighting matrix Ω, instrument Z, bandwidth constant c and moment function a.

Definition 5: The *kernel nonparametric efficiency bound* $\mathcal{M}(x_0, a)$ for estimating $\beta_0 = E_0\left(a|x_0\right)$ is the smallest possible value of $M(\Omega, Z, c, a)$ corresponding to the optimal choice of weighting matrix Ω, admissible instrument Z and bandwidth constant c.

Proposition 3. *Let Assumptions a.1, a.2 and A.1-A.25 in Appendix A.1 hold. (i) There exist an optimal weighting matrix $\Omega^*(a)$ and an optimal bandwidth constant $c^*(a)$. They are given in (A.9) and (A.10) in Appendix A.1.6. (ii) Any instrument:*

$$Z^* = E_0\left(\frac{\partial g'}{\partial \theta}\left(Y;\theta_0\right)|X\right) W\left(X\right), \tag{3.14}$$

where $W(X)$ is a positive definite matrix, P_0-a.s., is optimal, independent of a.

It is easily verified that $\mathcal{J}_{Z^*} = \mathcal{J}^*$ for any instrument Z^* in (3.14).

Corollary 4. *An admissible instrument Z is optimal if and only if the corresponding unconditional moment restrictions (3.10) identify all full-information identifiable parameters. The set of optimal instruments is independent of the selected kernel.*

Since we focus on the estimation of local conditional moment β_0, the set of optimal instruments is larger than the standard set of instruments for efficient estimation of a structural parameter θ_0 identified by (3.2). While in the standard framework $W(X) = V_0\left[g(Y;\theta_0)|X\right]^{-1}$ is the efficient weighting matrix for conditionally heteroskedastic moment restrictions [e.g., Chamberlain (1987)], any choice of a positive definite matrix $W(X)$ is asymptotically equivalent for estimating β_0 [13].

The expression of the kernel nonparametric efficiency bound is easily written in terms of the transformed parameters $(\eta_1^*, \eta_2^*)'$ defined in (3.5).

Proposition 5. *Under Assumptions a.1, a.2 and A.1-A.25 in Appendix A.1, the kernel nonparametric efficiency bound $\mathcal{M}(a, x_0)$ is given by the lower-right element of the matrix*

$$\left(J_0^{*\prime} \left(\frac{1}{c^{*d}} \frac{w^2}{f_X(x_0)} \Sigma_0 + c^{*2m} w_m^2 b(x_0) b(x_0)' \right)^{-1} J_0^* \right)^{-1}, \tag{3.15}$$

where $w^2 = \int_{\mathbb{R}^d} K(u)^2 du$, $w_m = \int_{\mathbb{R}} v^m \kappa(v) dv$, $c^* = c^*(a)$ is the optimal bandwidth constant given in (A.9) in Appendix A.1.6,

$$J_0^* = \begin{pmatrix} E_0\left(\frac{\partial g_2}{\partial \eta_2^{*\prime}} | x_0\right) & 0 \\ E_0\left(\frac{\partial a}{\partial \eta_2^{*\prime}} | x_0\right) & -1 \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} V_0(g_2|x_0) & Cov_0(g_2, a|x_0) \\ Cov_0(a, g_2|x_0) & V_0(a|x_0) \end{pmatrix},$$

and

$$b(x) = \frac{1}{m!} \frac{1}{f_X(x)} \begin{pmatrix} \Delta^m \left(E_0(g_2|x) f_X(x)\right) - E_0(g_2|x) \Delta^m f_X(x) \\ \Delta^m \left(E_0(a|x) f_X(x)\right) - E_0(a|x) \Delta^m f_X(x) \end{pmatrix},$$

with $\Delta^m := \sum_{i=1}^d \partial^m / \partial x_i^m$ and all functions evaluated at θ_0.

[13] In the application to derivative pricing, the state variables process is Markov. The instrument in (3.14) is optimal not only in the class of instruments written on X_t, but also in the class of instruments $Z_t = H(X_t, X_{t-1}, \cdots)$ written on the current and lagged values of the state variables, since the past is not informative once the present is taken into account.

The matrix in (3.15) resembles the GMM efficiency bound for estimating parameters $(\eta_2^*, \beta)'$ from orthogonality conditions based on function $\left(g_2', a-\beta\right)'$, with η_1^* known. Since moment restrictions (3.11)-(3.12) are conditional on $X = x_0$, the variance of the orthogonality conditions is replaced by the asymptotic MSE matrix $\frac{1}{c^{*d}} \frac{w^2}{f_X(x_0)} \Sigma_0 + c^{*2m} w_m^2 b(x_0) b(x_0)'$ of the kernel regression estimator at x_0 (up to a scale factor), and the expectations in the Jacobian matrix J_0^* are conditional on $X = x_0$. In particular, the efficiency bound depends on the likelihood of observing the conditioning variable close to x_0 by means of $f_X(x_0)$. The estimation of the full-information identified parameter η_1^* is irrelevant for the efficiency bound of β since the estimation of η_1^* achieves a faster parametric rate of convergence.

ii) Bias-free kernel nonparametric efficiency bound

If we restrict ourselves to asymptotically unbiased estimators for practical purposes, the bandwidth sequence has to be such that $\bar{c} = \lim T h_T^{2m+d} = 0$. When $\bar{c} = 0$ in Proposition 2, the asymptotic variance of $\sqrt{T h_T^d}\left(\widehat{\beta}_T - \beta_0\right)$ can be written as $w^2 V(Z, \Omega, a)$, where $V(Z, \Omega, a)$ is independent of the selected kernel K and bandwidth h_T.

Definition 6: The *bias-free kernel nonparametric efficiency bound* $\mathcal{B}(a, x_0)$ is the smallest asymptotic variance $V(Z, \Omega, a)$ corresponding to the optimal choice of the weighting matrix Ω and of the instrument Z.

Corollary 6. *(i) There exists an optimal choice of the weighting matrix Ω and of the instruments Z. The optimal instruments are given in (3.14) and the optimal weighting matrix in Appendix A.1.7. (ii) The bias-free kernel non-parametric efficiency bound $\mathcal{B}(x_0, a)$ is:*

$$
\begin{aligned}
\mathcal{B}(x_0, a) \;&= \frac{1}{f_X(x_0)} \left\{ V_0(a) - Cov_0(a, g_2) V_0(g_2)^{-1} Cov_0(g_2, a) \right. \\
&+ \left[E_0\left(\frac{\partial a}{\partial \theta'}\right) R_2 - Cov_0(a, g_2) V_0(g_2)^{-1} E_0\left(\frac{\partial g_2}{\partial \theta'}\right) R_2 \right] \left[R_2' E_0\left(\frac{\partial g_2'}{\partial \theta}\right) V_0(g_2)^{-1} E_0\left(\frac{\partial g_2}{\partial \theta'}\right) R_2 \right]^{-1} \\
&\left. \left[R_2' E_0\left(\frac{\partial a'}{\partial \theta}\right) - R_2' E_0\left(\frac{\partial g_2'}{\partial \theta}\right) V_0(g_2)^{-1} Cov_0(g_2, a) \right] \right\},
\end{aligned}
$$

where all moments are conditional on $X = x_0$ and evaluated at θ_0.

Since the expression of $\mathcal{B}(x_0, a)$ is a quadratic function of a, this formula holds for vector moment functions a, as well. When the parameter θ_0 itself is full-information identifiable, the bias-free kernel nonparametric efficiency bound becomes:

$$\mathcal{B}(x_0, a) = \frac{1}{f_X(x_0)} \left\{ V_0(a|x_0) - \text{ Cov}_0(a, g_2|x_0) V_0(g_2|x_0)^{-1} \text{ Cov}_0(g_2, a|x_0) \right\}. \quad (3.16)$$

Since the conditional moment of interest is also equal to $E_0\left(a|x_0\right) = E_0[a(Y;\theta_0) - Cov_0\left(a, g_2|x_0\right) V_0\left(g_2|x_0\right)^{-1} g_2(Y;\theta_0) \mid x_0]$, the bound (3.16) is simply the variance-covariance matrix of the residual term in the affine regression of a on g_2 performed conditional on x_0. A similar interpretation has already been given by Back and Brown (1993) in an unconditional setting [see also Brown and Newey (1998)], and extended to a conditional framework by Antoine, Bonnal and Renault (2007). In the general case, the efficiency bound $\mathcal{B}(x_0, a)$ balances the gain in information from the local conditional moment restrictions and the efficiency cost due to full-information underidentification of θ_0.

iii) Illustration with S&P 500 options

Let us derive the bias-free kernel nonparametric efficiency bounds for derivative prices estimation when the DGP P_0 is the stochastic volatility model (3.6)-(3.7) with parameters given by $\gamma_0 = 0.360$, $\rho_0 = 0.960$, $\delta_0 = 1.047$, $c_0 = 3.65 \cdot 10^{-6}$, $\theta_1^0 = .460 \cdot 10^{-6}$, $\theta_2^0 = -0.060$, $\theta_3^0 = 0.115$ and $\theta_4^0 = 0.860$. The riskfree term structure is as in Section 2.6 i)-ii). The ARG parameters ρ_0, δ_0, c_0 are set to match the stationary mean, variance and first-order autocorrelation of the realized volatility σ_t^2 of the S&P 500 index in the period from June, 1, 2001 to Mai, 31, 2005. The risk premia parameters θ_2^0 and θ_4^0 for stochastic volatility and underlying asset return, respectively, correspond to the XMM estimates obtained in Section 2.6 ii) for the S&P 500 options in June, 1, 2005 (see Table II). Parameters γ_0, θ_1^0, θ_3^0 are then fixed by the no-arbitrage restrictions (3.9). At a current date t_0, the prices of $n = 11$ actively traded call options are observed, with the same times-to-maturity and moneyness strikes as the S&P 500 put and call options with daily traded volume larger than 4000 contracts in June, 1, 2005 [see Section 2.6 iii)]. The current values of the state variables are the return and the realized volatility of the S&P 500 index on June, 1, 2005.

Let us focus on the time-to-maturity $h = 77$ days. The bias-free kernel nonparametric efficiency bound on the call option prices is displayed in Figure 3.

[Insert Figure 3: Bias-free kernel nonparametric efficiency bound, time-to-maturity 77-day]

The dashed line represents the theoretical call prices $E\left(a(k)|x_{t_0}\right)$ computed under the DGP P_0, and the dashed lines represents 95% asymptotic confidence intervals $E\left(a(k)|x_{t_0}\right) \pm 1.96\frac{w}{\sqrt{Th_T^2}}\mathcal{B}(x_{t_0}, k)^{1/2}$, as a function of moneyness k. The circles indicate the theoretical prices of the observed derivatives at time-to-maturity 77-day. In order to better visualize the pattern of the kernel nonparametric efficiency bound as a function of the moneyness strike, $\sqrt{w^2/Th_T^2}$ is set ten times larger than the value implied by the sample size and the bandwidths used in the empirical application. [14] The width of the confidence interval for derivative price $E\left(a(k)|x_{t_0}\right)$ depends on moneyness strike k. This width is zero when k corresponds to the moneyness strikes of the two observed calls at time-to-maturity $h = 77$. The width of the confidence interval is close to zero when the derivative is deep in-the-money, or deep out-of-the-money. Indeed, for moneyness strikes approaching zero or infinity, the kernel nonparametric efficiency bound goes to zero, since the option price has to be equal to the underlying asset price or equal to zero, respectively, by the no-arbitrage condition. [15] The confidence intervals in Figure 3 are generally larger for moneyness values $k < 1$ compared to $k > 1$. This is due to the different informational content of the set of observed option prices for the different moneyness strikes at the time-to-maturity of interest. Finally note that, when the confidence intervals and the bid-ask intervals don't intersect, some mispricing by the intermediaries, or the model, exists.

3.5 Asymptotic normality and efficiency of the XMM estimator

When the optimal instruments and optimal weighting matrix are used, the kernel moment estimator of β_0 in Definition 4 is kernel nonparametrically efficient. However, in application to derivative pricing this estimator does not ensure a positive estimated state price

[14]For expository purpose we consider symmetric confidence bands. These bands have to be truncated at zero to account for the positivity of derivative prices and get asymmetric bands. However, with the correct standardization, the truncation effect is negligible and arises only for large strikes. Moreover, the bands in Figure 3 are pointwise bands. Corollary 6 can be used to get ellipsoidal confidence sets for joint estimation of several derivative prices. It could also be possible to derive functional confidence sets for the entire pricing schedule as a function of moneyness strike k and for given time-to-maturity h, by considering $\hat{c}_{t_0}(h, k)$ as a stochastic process indexed by k. These developments are out of the scope of the paper.

[15]However, the relative accuracy can be poor in these moneyness regions.

density. This is a consequence of the quadratic GMM-type nature of the kernel moment estimators. The positivity of the estimated state price density is achieved by considering the information based XMM estimator defined in Section 2.5. In the general setting, the XMM estimator of β_0 is:

$$\hat{E}^*(a|x_0) = \int a(y;\hat{\theta})\hat{f}^*(y|x_0)dy,$$

where $\hat{\theta}$ is defined in (2.11)-(2.12) and $\hat{f}^*(y|x_0)$ is defined in (2.13).

The large sample properties of the XMM estimator of β_0 are given in Proposition 7 below, for a bandwidth sequence that eliminates the asymptotic bias.

Proposition 7. *Suppose the bandwidth is such that* $\bar{c} = \lim Th_T^{d+2m} = 0$. *Then the XMM estimator* $\hat{E}^*(a|x_0)$ *is consistent, converges at rate* $\sqrt{Th_T^d}$, *is asymptotically normal and bias-free kernel nonparametrically efficient:*

$$\frac{\sqrt{Th_T^d}}{w}(\hat{E}^*(a|x_0) - E_0(a|x_0)) \xrightarrow{d} N(0, \mathcal{B}(x_0, a)).$$

The XMM estimator of β_0 is asymptotically equivalent to the best kernel moment estimator of β_0 with optimal instruments and weighting matrix. By considering the constrained optimization of an information criterion (see Definition 1), the XMM estimator can be computed without making explicit the optimal instruments and weighting matrix.

The asymptotic distribution of the XMM estimator of θ_0 is given for the transformed parameter $\left(\eta_1^{*\prime}, \eta_2^{*\prime}\right)'$ in (3.5).

Corollary 8. *Suppose the bandwidth is such that* $\bar{c} = \lim Th_T^{d+2m} = 0$. *The XMM estimators* $\hat{\eta}_1^*$ *and* $\hat{\eta}_2^*$ *are asymptotically equivalent to the kernel moment estimators with optimal weighting matrix and instrument* Z *as in (3.14) with* $W(X) = V_0\left[g(Y;\theta_0)|X\right]^{-1}$.

In particular, the rates of convergence for the full-information identifiable parameter η_1^* and the full-information unidentifiable parameter η_2^* are $\sqrt{T}$ and $\sqrt{Th_T^d}$, respectively. The different rates of convergence of full-information identifiable and full-information unidentifiable parameters are reflected in the empirical results on the S&P 500 option data displayed in Table II in Section 2.6. If the DGP P_0 is the stochastic volatility model (3.6)-(3.7), parameter θ_4^0 is full-information identifiable (see Section 3.2) and its estimates are very stable over time, whereas the estimates of the full-information unidentifiable parameters θ_1^0, θ_2^0, θ_3^0 feature a larger time variability. Proposition 7 and Corollary 8 extend the first-order asymptotic equivalence between information based and quadratic GMM estimators [see Kitamura, Tripathi, Ahn (2004), Kitamura (2007)] to a setting including local conditional moment restrictions and allowing for full-information unidentifiable parameters.

4 Concluding remarks

The literature on joint estimation of historical and risk-neutral parameters is generally based on either Maximum Likelihood (ML), or GMM, type of methods. A part of this literature relies on uniform moment restrictions from a time series of spot prices only, and implicitly assumes that the risk premia parameters are full-information identified [e.g., Bansal, Viswanathan (1993), Hansen, Jagannathan (1997), Stock, Wright (2000)]. This assumption is not valid when some risk premia parameters can be identified only from option data. Another part of the literature exploits time series of both spot and option prices [e.g., Duan (1994), Chernov, Ghysels (2000), Pan (2002), Eraker (2004)]. However the activity on derivative markets is rather weak, and these approaches typically rely either on artificial option series which are approximately near-the-money and at short time-to-maturity, or on option quotes of both actively and less actively traded options, or on ad-hoc assumptions on time-varying options characteristics.

In this paper we introduce a new XMM estimator of derivative prices using jointly a time series of spot returns and a few cross-sections of derivative prices. We argue that these two types of data imply different types of conditional moment restrictions, that are uniform and local, respectively. First, the XMM approach allows for consistent estimation of the sdf parameters θ even if they are full-information unidentifiable. Second, the XMM estimator of the pricing operator at a given date is consistent for a fixed number of cross-sectionally observed derivative prices. Third, the XMM estimator is asymptotically efficient. These results are due to both the parametric sdf and the deterministic relationships between derivative prices that hold in a no-arbitrage pricing model with a finite number of state variables. The application to the S&P 500 options shows that the new XMM-based calibration approach outperforms the traditional cross-sectional calibration approach while being easy to implement. In particular, the XMM estimated option prices are compatible with the observed option prices for highly traded derivatives, and are more stable over time. Finally, the asymptotic results that have been derived for the XMM estimator can be used to develop test procedures. Tests of correct specification of the parametric sdf can be based on the minimized XMM objective function. Overidentification tests could also be defined by increasing the number of local restrictions, that is the number of observed option prices in the application to derivative pricing, and comparing the XMM estimators with and without additional local restrictions. This would extend the standard overidentification tests introduced by Hansen (1982) and Szroeter (1983) for full-information identifiable parameters.

The XMM approach can be applied to other financial markets with heterogenous trading activity. The T-bond market, for instance, has similar features as the index derivatives market: regular issuing of standardized products, and small number of times-to-maturity which are highly traded daily. The XMM approach allows for a first correction of alternative estimation and pricing methods that assume a high activity for all assets proposed on the markets. It will have to be completed by a model describing the activity on financial markets, which is a challenging topic for future research.

References

[1] Ai, C., and X. Chen (2003): Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions, Econometrica, 71, 1795-1843.

[2] Aït-Sahalia, Y., and J. Duarte (2003): Nonparametric Option Pricing under Shape Restrictions, Journal of Econometrics, 116, 9-47.

[3] Aït-Sahalia, Y., and A. Lo (1998): Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Returns, Journal of Finance, 53, 499-547.

[4] Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003): Modeling and Forecasting Realized Volatility, Econometrica, 71, 529-626.

[5] Andrews, D. (1991): An Empirical Process Central Limit Theorem for Dependent Non-identically Distributed Random Variables, Journal of Multivariate Analysis, 38, 187–203.

[6] Andrews, D. (1994): Empirical Process Methods in Econometrics, in Handbook of Econometrics, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: North Holland, 2247-2294.

[7] Antoine, B., H. Bonnal, and E. Renault (2007): On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood, Journal of Econometrics, 138, 461-487.

[8] Back, K., and D. Brown (1992): GMM, Maximum Likelihood and Nonparametric Efficiency, Economics Letters, 39, 23-28.

[9] Back, K., and D. Brown (1993): Implied Probabilities in GMM Estimators, Econometrica, 61, 971-975.

[10] Bakshi, G., C. Cao, and Z. Chen (1997): Empirical Performance of Alternative Option Pricing Models, Journal of Finance, 52, 2003-2049.

[11] Bansal, R., and S. Viswanathan (1993): No Arbitrage and Arbitrage Pricing: A New Approach, Journal of Finance, 48, 1231-1362.

[12] Billingsley, P. (1965): Ergodic Theory and Information. New York: John Wiley.

[13] Bondarenko, O. (2003): Estimation of Risk-neutral Densities using Positive Convolution Approximation, Journal of Econometrics, 116, 85-112.

[14] Bosq, D. (1998): Nonparametric Statistics for Stochastic Processes: Estimation and Prediction. New York: Springer.

[15] Brown, B., and W. Newey (1998): Efficient Semiparametric Estimation of Expectations, Econometrica, 66, 453-464.

[16] Buchen, P., and M. Kelly (1996): The Maximum Entropy Distribution of an Asset Inferred from Option Prices, Journal of Financial and Quantitative Analysis, 31, 143-159.

[17] Carr, P., and D. Madan (1999): Option Pricing and the Fast Fourier Transform, Journal of Computational Finance, 2, 61-73.

[18] Chamberlain, G. (1987): Asymptotic Efficiency in Estimation with Conditional Moment Restrictions, Journal of Econometrics, 34, 305-334.

[19] Chernov, M., and E. Ghysels (2000): A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Option Valuation, Journal of Financial Economics, 56, 407-458.

[20] Cox, J., J. Ingersoll, and S. Ross (1985): A Theory of the Term Structure of Interest Rates, Econometrica, 53, 385-407.

[21] Darolles, S., C. Gouriéroux, and J. Jasiak (2006): Structural Laplace Transform and Compound Autoregressive Models, Journal of Time Series Analysis, 27, 477-504.

[22] Duan, J. (1994): Maximum Likelihood Estimation Using Price Data of the Derivative Contract, Mathematical Finance, 4, 155-167.

[23] Eraker, B. (2004): Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, Journal of Finance, 59, 1367-1403.

[24] Gagliardini, P., C. Gouriéroux, and E. Renault (2010): ”Efficient Derivative Pricing by the Extended Method of Moments, Supplementary material: Appendix B”, http://www.people.usi.ch/gagliarp/AppendixB.pdf.

[25] Gouriéroux, C., and J. Jasiak (2001): Financial Econometrics: Problems, Models and Methods. Princeton, NJ: Princeton University Press.

[26] Gouriéroux, C., and J. Jasiak (2006): Autoregressive Gamma Process, Journal of Forecasting, 25, 129-152.

[27] Hansen, L. (1982): Large Sample Properties of Generalized Method of Moments Estimators, Econometrica, 50, 1029-1054.

[28] Hansen, L., and R. Jagannathan (1997): Assessing Specification Errors in Stochastic Discount Factor Models, Journal of Finance, 52, 557-590.

[29] Hansen, L., and S. Richard (1987): The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models, Econometrica, 55, 587-613.

[30] Hansen, L., and K. Singleton (1982): Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models, Econometrica, 50, 1269-1286.

[31] Herrndorf, N. (1984): A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables, Annals of Probability, 12, 141-153.

[32] Heston, S. (1993): A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, Review of Financial Studies, 6, 327-343.

[33] Hull, J. (2005): Options, Futures and Other Derivatives (Seventh Ed.). Upper Saddle River, NJ: Prentice Hall.

[34] Hull, J., and A. White (1987): The Pricing of Options on Assets with Stochastic Volatilities, Journal of Finance, 42, 281-300.

[35] Imbens, G., R. Spady, and P. Johnson (1998): Information Theoretic Approaches to Inference in Moment Conditions Models, Econometrica, 66, 333-357.

[36] Jackwerth, J. (2000): Recovering Risk Aversion from Option Prices and Realized Returns, Review of Financial Studies, 13, 433-451.

[37] Jackwerth, J., and M. Rubinstein (1996): Recovering Probability Distributions from Option Prices, Journal of Finance, 51, 1611-1631.

[38] Kitamura, Y. (2007): Empirical Likelihood Methods in Econometrics: Theory and Practice, in Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society, Vol. 3, ed. by R. Blundell, W. Newey and T. Persson. Cambridge, U.S.: Cambridge University Press, 174-237.

[39] Kitamura, Y., and M. Stutzer (1997): An Information Theoretic Alternative to Generalized Method of Moments Estimation, Econometrica, 65, 861-874.

[40] Kitamura, Y., G. Tripathi, and H. Ahn (2004): Empirical Likelihood Based Inference in Conditional Moment Restriction Models, Econometrica, 72, 1667-1714.

[41] Pan, J. (2002): The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time Series Study, Journal of Financial Economics, 63-50.

[42] Pollard, D. (1990): Empirical Processes: Theory and Applications, CBMS Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: IMS.

[43] Potscher, B., and I. Prucha (1989): A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes, Econometrica, 57, 675-683.

[44] Schwartz, E. (1987): The Stochastic Behaviour of Commodity Prices: Implications for Valuation and Hedging, Journal of Finance, 52, 923-973.

[45] Silverman, B. (1986): Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

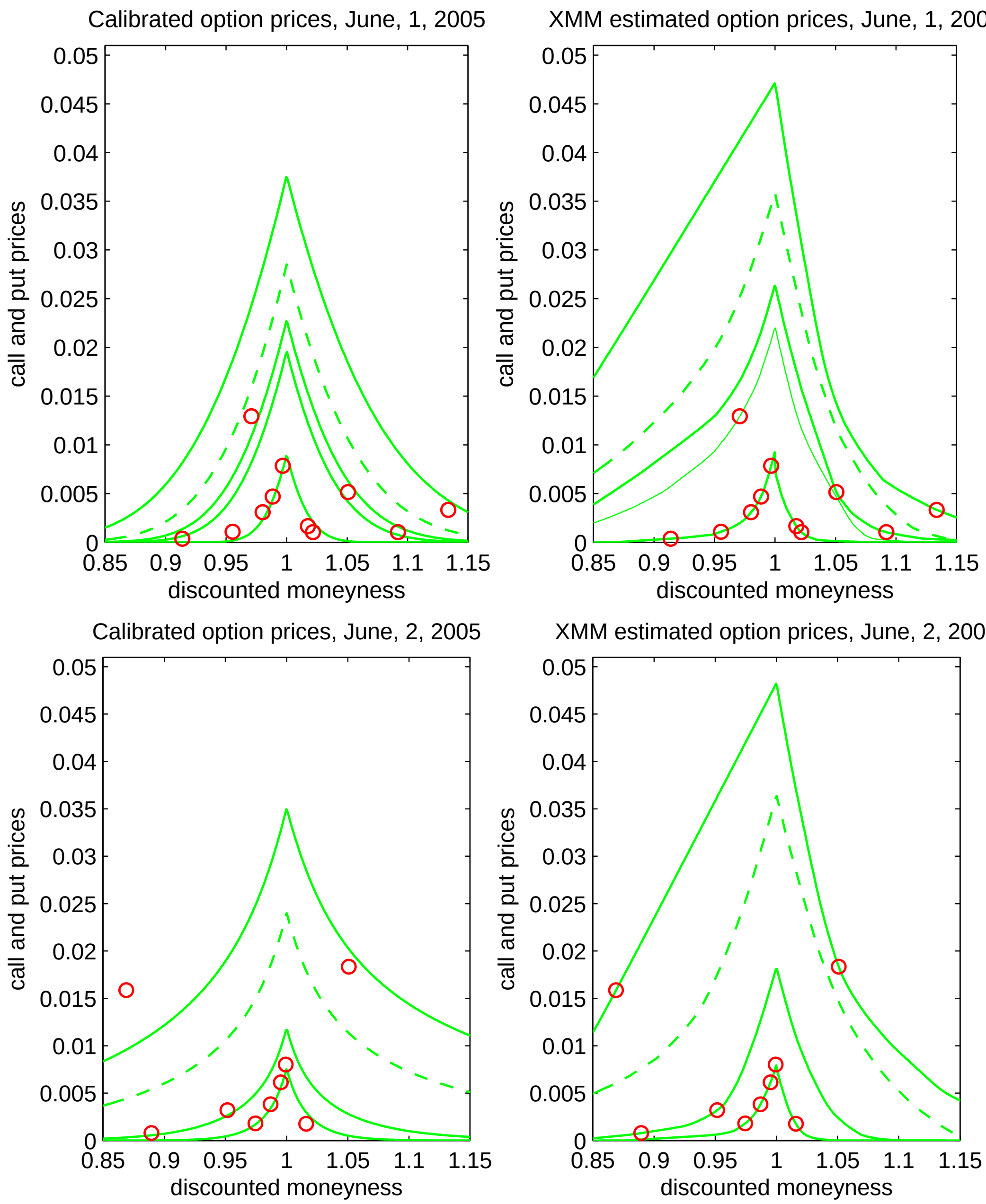
[46] Stock, J., and J. Wright (2000): GMM with Weak Identification, Econometrica, 68, 1055-1096.

[47] Stutzer, M. (1996): A Simple Nonparametric Approach to Derivative Security Valuation, Journal of Finance, 51, 1633-1652.

[48] Szroeter, J. (1983): Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions, Econometrica, 51, 335-353.

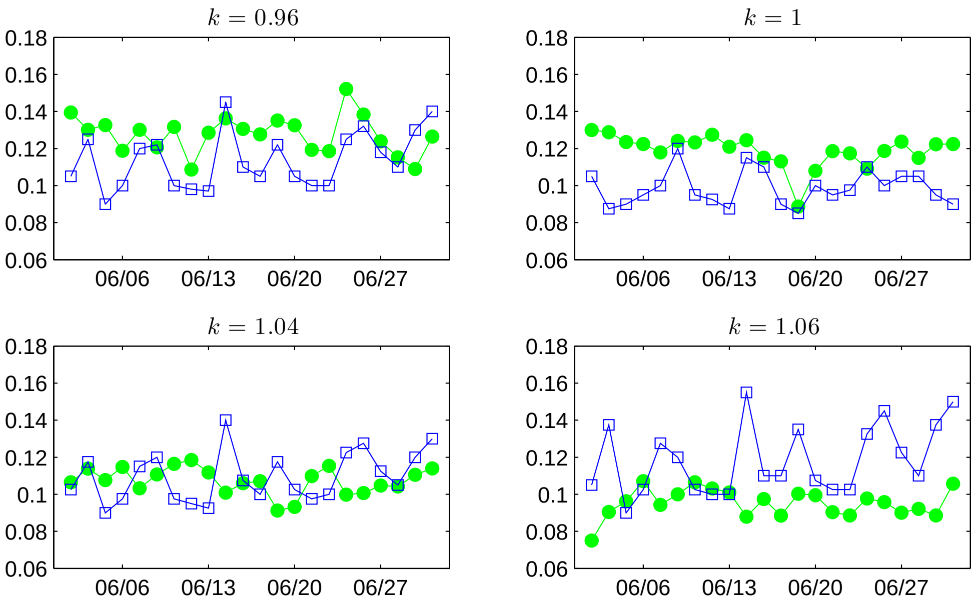
[49] Tenreiro, C. (1995): Estimation fonctionnelle: Applications aux tests d'adéquation et de paramètre constant, Ph.D. Thesis, University of Paris VI.

Figure 1: Estimated call and put prices for S&P 500 options at June, 1, and June, 2, 2005.



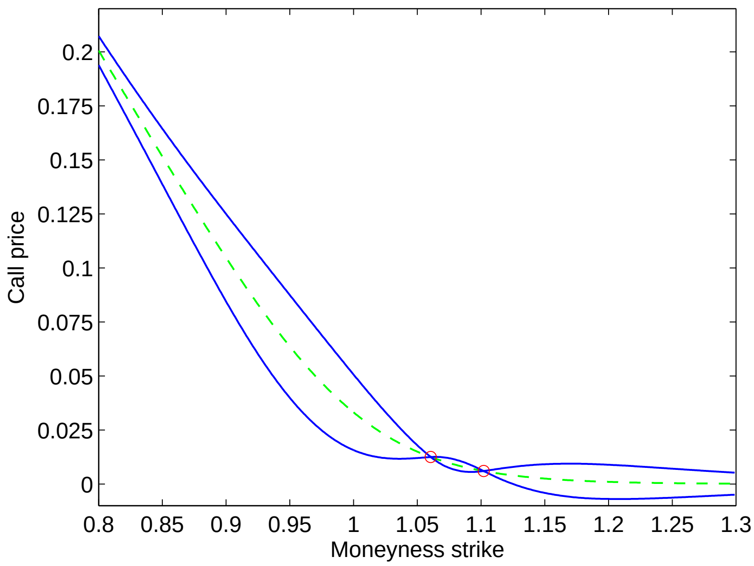
In the upper right Panel, the solid lines correspond to estimated relative option prices as a function of discounted moneyness strike $B(t, t+h)k$ for the highly traded times-to-maturity $h = 12, 57, 77, 209$ at June, 1, 2005, obtained by XMM. The dashed line corresponds to XMM estimated prices for the non-traded time-to-maturity $h = 120$. The price curves correspond to puts if $B(t, t+h)k < 1$, to calls otherwise. In the upper left Panel, the solid and dashed lines are the price curves obtained by the parametric pricing model (2.14)-(2.15) with the calibrated parameters in Table 1 for times-to-maturity $h = 12, 57, 77, 209$, and $h = 120$, respectively. In both Panels, circles correspond to observed S&P 500 option prices with daily trading volume larger than 4000 contracts. The two lower Panels correspond to June, 2, 2005 with highly traded times-to-maturity $h = 11, 31, 208$, and non-traded time-to-maturity $h = 119$.

Figure 2: Time series of implied volatilities of S&P 500 options in June, 2005.



In Panels 1-4, annualized implied volatilities at time-to-maturity $h = 20$ and moneyness strike $k = .96, 1, 1.04, 1.06$, respectively, are displayed for each trading day in June 2005. Circles are implied volatilities computed from option prices estimated by the XMM approach, squares are implied volatilities from the cross-sectional calibration approach. The ticks on the horizontal axis correspond to Mondays.

Figure 3: Bias-free kernel nonparametric efficiency bound at June, 1, 2005, time-to-maturity 77.



The dashed line corresponds to the relative call price $E\left(a(k)|x_{t_0}\right)$ at time-to-maturity $h = 77$, the solid lines to pointwise 95% symmetric confidence intervals $E\left(a(k)|x_{t_0}\right) \pm 1.96\frac{w}{\sqrt{Th_T^2}}\mathcal{B}(x_{t_0}, k)^{1/2}$. The value of $\sqrt{w^2/Th_T^2}$ is set 10 times larger than in the empirical application.

Table I: Calibrated parameters (cross-sectional approach) for the S&P 500 options in June, 2005.

Day	Calibrated parameters				Goodness of fit
t_0	$\hat{\delta}^*_{t_0}$	$\hat{\rho}^*_{t_0}$	$\hat{c}^*_{t_0}$ $(\times 10^{-7})$	$\hat{\sigma}_{t_0}$	$RMSE_{t_0}$
1.6.05	24.7303	0.47588	9.59115	0.00347	0.00161
2.6.05	0.05400	0.99994	94.2208	0.00670	0.00219
3.6.05	87.7199	0.99999	0.09906	0.00480	0.00114
6.6.05	1.68952	0.99625	6.57284	0.00519	0.00194
7.6.05	0.41250	0.99998	41.2201	0.00563	0.00071
8.6.05	10.8766	0.81153	12.5623	0.00102	0.00061
9.6.05	2.26166	0.99106	6.58317	0.00501	0.00191
10.6.05	3.15278	0.96637	6.79999	0.00499	0.00032
13.6.05	0.21085	0.99909	8.84000	0.00555	0.00035
14.6.05	0.02125	0.99999	101.489	0.00874	0.00267

Calibrated parameter $\hat{\theta}_{t_0}$, volatility $\hat{\sigma}_{t_0}$ and goodness of fit measure $RMSE_{t_0}$ for the first ten trading days t_0 of June 2005. The calibration is performed using a Fourier Transform approach to compute option prices. At each day t_0, the sample consists of the derivative prices at t_0 of S&P 500 options with daily volume larger than 4000 contracts.

Table II: Estimated sdf parameters and option prices (XMM approach) for the S&P 500 options in June, 2005.

Day	Sdf parameters				Option prices $(\times 10^{-2})$				
t_0	$\hat{\theta}_1$ $(\times 10^{-6})$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$k = .96$	$k = .98$	$k = 1$	$k = 1.02$	$k = 1.04$
1.6.05	.910	$-.059$	.121	.860	.269	.631	1.356	.651	.162
2.6.05	.897	$-.057$	.118	.860	.215	.584	1.342	.686	.204
3.6.05	.900	$-.054$	.115	.860	.229	.539	1.282	.637	.169
6.6.05	.892	$-.056$	.118	.860	.155	.496	1.270	.622	.210
7.6.05	.888	$-.056$	.117	.860	.218	.550	1.227	.514	.147
8.6.05	.891	$-.063$	.124	.860	.164	.531	1.288	.628	.186
9.6.05	.890	$-.063$	.124	.860	.223	.574	1.280	.607	.220
10.6.05	.878	$-.063$	.125	.860	.110	.495	1.326	.701	.233
13.6.05	.864	$-.063$	.124	.860	.206	.553	1.253	.593	.193
14.6.05	.851	$-.063$	.125	.860	.250	.588	1.295	.580	.133

Estimated sdf parameter $\hat{\theta}$ and relative option prices $\hat{c}_{t_0}(h, k)$ at time-to-maturity $h = 20$ for the first ten trading days t_0 of June 2005. The option prices correspond to puts for $k < 1$, and to calls for $k \geq 1$. The estimation is performed using XMM. At each day t_0, the sample consists of the current and previous $T = 1000$ observations on the state variables, and the derivative prices at t_0 of S&P 500 options with daily volume larger than 4000 contracts.

APPENDIX A: Proofs

In this Appendix we give the proofs of Propositions 2, 3 and 5 (Section A.1), Proposition 7 (Section A.2), and Proposition 1, as well as of other results of Section 3.2 concerning the parametric stochastic volatility model (Section A.3).

A.1 Asymptotic properties of kernel moment estimators

We use the following notation. Symbol $\Longrightarrow$ denotes weak convergence in the space of bounded real functions on set $\Theta \subset \mathbb{R}^p$, equipped with the uniform metric [see e.g. Andrews (1994)]. The Frobenius norm of matrix A is $\|A\| = \left[Tr\left(AA^{'}\right)\right]^{1/2}$. For a multi-index $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}^d$ and vector $x \in \mathbb{R}^d$, we set $|\alpha| := \sum_{i=1}^d \alpha_i$, $x^\alpha := x_1^{\alpha_1} \cdot ... \cdot x_d^{\alpha_d}$, and $\partial^{|\alpha|} f / \partial x^\alpha := \partial^{|\alpha|} f / \partial x_1^{\alpha_1} ... \partial x_d^{\alpha_d}$. Symbol $\|f\|_\infty$ denotes the sup-norm $\|f\|_\infty = \sup_{x \in \mathcal{X}} \|f(x)\|$ of a continuous function f defined on set $\mathcal{X}$. We denote by $C^m(\mathcal{X})$ the space of functions f on $\mathcal{X}$ that are continuously differentiable up to order $m \in \mathbb{N}$, $\|D^m f\|_\infty := \sum_{|\alpha|=m} \left\|\partial^{|\alpha|} f / \partial x^\alpha\right\|_\infty$, and $\Delta^m f := \sum_{i=1}^d \partial^m f / \partial x_i^m$. Furthermore, $L^2(F_Y)$ denotes the Hilbert space of real-valued functions, which are square integrable w.r.t. the distribution F_Y of r.v. Y, and $\|.\|_{L^2(F_Y)}$ is the corresponding L^2-norm. Linear space $L^p(\mathcal{X})$, $p > 0$, of p-integrable functions w.r.t. Lebesgue measure on set $\mathcal{X}$ is defined similarly. We denote by g_2^* the function defined by $g_2^*(y;\theta) = \left(g_2(y;\theta)^{'}, a(y;\theta)^{'}\right)^{'}$. Finally, all functions of θ are evaluated at θ_0, when the argument is not explicit, and the expectation $E[.]$ is w.r.t. the DGP P_0.

A.1.1 Regularity assumptions

Let us introduce the following set of regularity conditions:

Assumption A.1: *The instrument Z is given by $Z = H(X)$, where H is a matrix function defined on $\mathcal{X}$ and is continuous at $x = x_0$.*

Assumption A.2: *The true value of the parameter $\theta_0 \in \mathbb{R}^p$ is globally identified with instrument Z, that is,*

$$\left(E\left[g_1(X,Y;\theta)\right]^{'}, E\left[g_2(Y;\theta) \mid X = x_0\right]^{'}\right)^{'} = 0, \, \theta \in \Theta \quad \Longrightarrow \quad \theta = \theta_0,$$

where $g_1(X,Y;\theta) = Z \cdot g(Y;\theta)$.

Assumption A.3: *The true value of the parameter θ_0 is locally identified with instrument Z, that is, the matrix*

$$\begin{pmatrix} E\left[\frac{\partial g_1}{\partial \theta'}(X,Y;\theta_0)\right] \\ E\left[\frac{\partial g_2}{\partial \theta'}(Y;\theta_0) \mid X = x_0\right] \end{pmatrix}$$

has full column-rank.

Assumption A.4: *The parameter sets $\Theta \subset \mathbb{R}^p$ and $B \subset \mathbb{R}^L$ are compact and the true parameter $\theta_0^* = \left(\theta_0^{'}, \beta_0^{'}\right)^{'}$ is in the interior of $\Theta \times B$, where $\beta_0 = E\left[a(Y;\theta_0)|X = x_0\right]$.*

Assumption A.5: *The process $\left\{\left(X_t^{'}, Y_t^{'}\right)^{'} : t \in \mathbb{N}\right\}$ on $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^d \times \mathbb{R}^{\tilde{d}}$ is strictly stationary and geometrically strongly mixing.*

Assumption A.6: *The stationary density f_X of X_t is in class $C^m(\mathcal{X})$ for some $m \in \mathbb{N}$, $m \geq 2$, and is such that $\|f_X\|_\infty < \infty$ and $\|D^m f_X\|_\infty < \infty$.*

Assumption A.7: *For $t_1 < t_2$, the stationary density f_{t_1,t_2} of (X_{t_1}, X_{t_2}) is such that $\sup_{t_1<t_2} \|f_{t_1,t_2}\|_\infty < \infty$. Moreover, for $t_1 < t_2 < t_3 < t_4$, the stationary density f_{t_1,t_2,t_3,t_4} of $(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})$ is such that:*

$$\sup_{t_1<t_2<t_3<t_4} \|f_{t_1,t_2,t_3,t_4}\|_\infty < \infty.$$

Assumption A.8: *The product kernel $K(u) = \prod_{i=1}^{d} \kappa(u_i)$, $u \in \mathbb{R}^d$, is such that: i) $\int_{\mathbb{R}^d} K(u)du = 1$, ii) K is bounded, $\lim_{\|u\|\to\infty} \|u\|^d K(u) = 0$, $\int_{\mathbb{R}^d} |K(u)|\, du < \infty$ and $w^2 := \int_{\mathbb{R}^d} K(u)^2 du < \infty$, iii) $\int_{\mathbb{R}} v^l \kappa(v)dv = 0$ for any $l \in \mathbb{N}$ such that $l < m$, and $\int_{\mathbb{R}} \kappa(v)\, |v|^m\, dv < \infty$.*

Assumption A.9: *The bandwidth h_T is such that $T^{1/2}h_T^d \to \infty$ and $Th_T^{d+2m} \to \bar{c} \in [0,\infty)$, as $T \to \infty$.*

Assumption A.10: *The matrices:*

$$S_0 = V\left[g_1\left(X_t, Y_t; \theta_0\right)\right], \qquad \Sigma_0 = V\left[g_2^*\left(Y_t; \theta_0\right) \middle| X_t = x_0\right],$$

exist and are positive definite.

Assumption A.11: *For any $\theta \in \Theta$: $E\left[\|g_1\left(X_t, Y_t; \theta\right)\|^4\right] < \infty$, $E\left[\|g_2^*\left(Y_t; \theta\right)\|^4\right] < \infty$.*

Assumption A.12: *For any $\theta \in \Theta$, the function $x \longmapsto \varphi(x;\theta) = E\left[g_2^*(Y_t;\theta)|X_t = x\right] f_X(x)$ is in class $C^m(\mathcal{X})$, such that $\sup_{\theta\in\Theta} \|D^m\varphi(.;\theta)\|_\infty < \infty$ and $\partial^{|\alpha|}\varphi/\partial x^\alpha$ is uniformly continuous on $\mathcal{X} \times \Theta$ for any $\alpha \in \mathbb{N}^d$ with $|\alpha| = m$.*

Assumption A.13: *For any $\theta, \tau \in \Theta$, the function $E\left[g_2^*(Y_t;\theta)g_2^*(Y_t;\tau)'|X_t = .\right] f_X(.)$ is continuous at $x = x_0$.*

Assumption A.14: *For any $\theta \in \Theta$, $\sup_{t_1<t_2} \|E\left[\|g_2^*(Y_{t_1};\theta)\|^2|X_{t_1} = ., X_{t_2} = .\right] f_{t_1,t_2}(.,.)\,\|_\infty < \infty$.*

Assumption A.15: *For any $\theta \in \Theta$,*

$$\sup_{t_1\leq t_2\leq t_3\leq t_4} \|E\left[\,\|g_2^*(Y_{t_1};\theta)\|\,\|g_2^*(Y_{t_2};\theta)\|\,\|g_2^*(Y_{t_3};\theta)\|\,\|g_2^*(Y_{t_4};\theta)\|\right.$$
$$\left.|\; X_{t_1} = ., X_{t_2} = ., X_{t_3} = ., X_{t_4} = .\right] f_{t_1,t_2,t_3,t_4}(.,.,.,.)\|_\infty < \infty.$$

Assumption A.16: *There exists a basis of functions $\left\{\psi_j : j \in \mathbb{N}\right\}$ in $L^2(F_Y)$, where F_Y is the stationary cdf of Y_t, such that $\left\|\psi_j\right\|_{L^2(F_Y)} = 1$, $j \in \mathbb{N}$, and:*

$$g_2^*(y;\theta) = \sum_{j=1}^{\infty} c_j(\theta)\, \psi_j(y), \qquad y \in \mathcal{Y},$$

for any $\theta \in \Theta$, where $\{c_j(\theta) : j \in N\}$ is a sequence of coefficient vectors. Moreover, there exist $r > 2$ and a sequence $\{\lambda_j > 0 : j \in \mathbb{N}\}$, such that $\sum_{j=1}^{\infty} \lambda_j < \infty$, and:

$$\sum_{j=1}^{\infty} \lambda_j \left(E\left[\left\|Z_t\psi_j(Y_t)\right\|^r\right]^{2/r} + E\left[\psi_j\left(Y_t\right)^2 |X_t = x_0\right]\right) < \infty, \qquad \lim_{J\to\infty} \sup_{\theta\in\Theta} \sum_{j=J}^{\infty} \frac{1}{\lambda_j} \left\|c_j(\theta)\right\|^2 = 0.$$

Assumption A.17: *The function* $x \longmapsto \varphi_j(x) = E\left[\psi_j(Y_t)^2 | X_t = x\right] f_X(x)$ *is in class* $C^2(\mathcal{X})$, *for any* $j \in \mathbb{N}$, *such that* $\sup_{j \in \mathbb{N}} \left\|\varphi_j\right\|_\infty < \infty$ *and* $\sup_{j \in \mathbb{N}} \left\|D^2 \varphi_j\right\|_\infty < \infty$.

Assumption A.18: *The functions* ψ_j *are such that* $\sup_{j \in \mathbb{N}} E\left[\left|\psi_j(Y_t)\right|^{\bar{r}}\right] < \infty$, *for* $\bar{r} > 2$, *and:*

$$\sup_{j \in \mathbb{N}} \sup_{t_1 < t_2} \left\| E\left[\psi_j(Y_{t_1}) \psi_j(Y_{t_2}) \| X_{t_1} = ., X_{t_2} = .\right] f_{t_1, t_2}(., .) \right\|_\infty < \infty.$$

Assumption A.19: *The moment function* $\theta \longmapsto \left(E\left[g_1\left(X_t, Y_t; \theta\right)\right]', E\left[g_2^*\left(Y_t; \theta\right) | X_t = x_0\right]'\right)'$ *is continuous on* Θ.

Assumption A.20: *The weighting matrix* Ω *is positive definite.*

Assumption A.21: *Function* $g_2^*(y; \theta)$ *is twice continuously differentiable w.r.t.* $(y, \theta) \in \mathcal{Y} \times \Theta$.

Assumption A.22: *There exist* $\gamma_1, \gamma_2 > 1$ *and* $\tau > 2$, *such that:*

$$E\left[\left\|\frac{\partial g_1}{\partial \theta'}\left(X_t, Y_t; \theta_0\right)\right\|^\tau\right] < \infty \quad , \quad E\left[\sup_{\theta \in \Theta}\left\|\frac{\partial g_1}{\partial \theta'}\left(X_t, Y_t; \theta\right)\right\|^{\gamma_1}\right] < \infty,$$
$$E\left[\sup_{\theta \in \Theta}\left\|\frac{\partial^2 g_1}{\partial \theta_i \partial \theta_j}\left(X_t, Y_t; \theta\right)\right\|^{\gamma_2}\right] < \infty \quad , \quad i, j = 1, ..., p.$$

Assumption A.23: *The function* $E\left[\frac{\partial g_2^*}{\partial \theta'}(Y_t; \theta_0) | X_t = .\right] f_X(.)$ *is continuous at* x_0, *and:*

$$E\left[\left\|\frac{\partial g_2^*}{\partial \theta'}(Y_t; \theta_0)\right\|^\delta\right] < \infty,$$

for $\delta > 2$. *Moreover, the functions:*

$$E\left[\sup_{\theta \in \Theta}\left\|\frac{\partial^2 g_2^*}{\partial \theta_i \partial \theta_j}(Y_t; \theta)\right\| | X_t = .\right] f_X(.), \quad i, j = 1, ..., p,$$

are bounded.

Assumption A.24: *The functions:*

$$\theta \longmapsto \left(E\left[\frac{\partial g_1'}{\partial \theta}\left(X_t, Y_t; \theta\right)\right], E\left[\frac{\partial g_2^{*\prime}}{\partial \theta}\left(Y_t; \theta\right) | X_t = x_0\right]\right),$$
$$\theta \longmapsto E\left[\frac{\partial^2 g_1}{\partial \theta_i \partial \theta_j}\left(X_t, Y_t; \theta\right)\right], \qquad i, j = 1, ..., p,$$

are continuous on Θ.

Assumption A.25: *The matrix* $Z^* = E\left(\frac{\partial g'}{\partial \theta}\left(Y; \theta_0\right) | X\right) W(X)$, *where* $W(X)$ *is a positive definite matrix* P_0*-a.s., is an instrument satisfying Assumptions A.1, A.10, A.11, A.16, A.19, A.22 and A.24.*

Due to the different rates of convergence of the empirical moments in $\widehat{g}_T(\theta^*)$ in Definition 4, it is not possible to follow the standard approach for the GMM framework to derive the asymptotic properties of $\widehat{\theta}_T^*$. In particular, to prove consistency, we cannot rely on the uniform a.s. convergence of the criterion Q_T to a limit deterministic function. Indeed, after dividing Q_T by T, the part of the criterion involving the local conditional moment restrictions is asymptotically negligible. The corresponding limiting criterion is not uniquely minimized at θ_0 in the full-information underidentified case. To prove consistency and asymptotic normality of $\widehat{\theta}_T^*$, we follow an alternative approach relying on empirical process methods [see Stock, Wright (2000) for a similar approach].

The empirical process of the sample moment restrictions is:

$$\Psi_T(\theta) = \widehat{g}_T(\theta^*) - m_T(\theta^*) =: T^{-1/2} \sum_{t=1}^{T} g_{t,T}(\theta), \quad \theta \in \Theta, \tag{A.1}$$

where:

$$m_T(\theta^*) = \left(\sqrt{T} E\left[g_1(X,Y;\theta)\right]', \sqrt{Th_T^d} E\left[g_2(Y;\theta)|x_0\right]', \sqrt{Th_T^d} E\left[a(Y;\theta) - \beta|x_0\right]' \right)'.$$

Due to the linearity of $\widehat{g}_T$ and m_T w.r.t. β, the empirical process Ψ_T is function of parameter θ, but not of parameter β. Moreover, the triangular array $g_{tT}(\theta)$ is not zero-mean, because of the bias term in the nonparametric component. Assumptions A.1 and A.4-A.19 are used to prove the weak convergence of the empirical process Ψ_T to a Gaussian stochastic process on Θ (see Lemma A.1 in Section A.1.2). In particular, Assumption A.5 on weak serial dependence, Assumptions A.6-A.7 on the smoothness of the stationary density of (X_t), Assumptions A.8 and A.9 on the kernel and the bandwidth, and Assumptions A.10-A.15 on the moment functions, yield the asymptotic normality of the finite-dimensional distributions of process Ψ_T. Assumption A.16 disentangles the dependence of $g_2^*(y;\theta)$ on y and θ. It is used, together with Assumptions A.17 and A.18, to prove the stochastic equicontinuity of process Ψ_T along the lines of Andrews (1991). The weak convergence of empirical process Ψ_T is combined with Assumption A.2 on global identification, Assumption A.19 on the continuity of the moment functions and Assumption A.20 on the weighting matrix, to prove the consistency of kernel moment estimator $\widehat{\theta}_T^*$ (see Section A.1.3).

Assumptions A.21-A.24 concern the first- and second-order derivatives of the moment functions w.r.t. the parameter θ. These assumptions are used, together with the local identification Assumption A.3 and Assumption A.4 of interior true parameter value, to derive an asymptotic expansion for $\widehat{\theta}_T^*$ in terms of $\Psi_T(\theta_0)$. From the asymptotic normality of $\Psi_T(\theta_0)$ (Lemma A.1) we deduce the asymptotic normality of kernel moment estimator $\widehat{\theta}_T^*$ (see Section A.1.4). Finally, Assumption A.25 is used to establish the asymptotic results for the kernel moment estimators with optimal instruments (see Section A.1.6).

Let us now discuss the bandwidth conditions in Assumption A.9. The condition $Th_T^{d+2m} \to \bar{c} \in [0, \infty)$ is standard in nonparametric regression analysis. When $\bar{c} > 0$, the bandwidth features the optimal d-dimensional rate of convergence, whereas when $\bar{c} = 0$ the asymptotic bias becomes negligible. Condition $T^{1/2}h_T^d \to \infty$ is stronger than the standard condition $Th_T^d \to \infty$. Such a stronger bandwidth condition is necessary to ensure negligible second-order terms in the asymptotic expansion of the kernel moment estimator. Indeed, in the full-information underidentified case, some linear combinations of parameter θ_0 are estimated at a nonparametric rate $\sqrt{Th_T^d}$, whereas other linear combinations are estimated at a parametric rate $\sqrt{T}$. Thus, we need to ensure that the second-order term with smallest rate of convergence is negligible w.r.t. the first-order term with largest rate of convergence:

$$\left(1/\sqrt{Th_T^d}\right)^2 = o(1/\sqrt{T}) \Longleftrightarrow T^{1/2}h_T^d \to \infty.$$

The bandwidth conditions in Assumption A.9 can be satisfied when $d < 2m$. In particular, a kernel of order $m = 2$ is sufficient when the dimension $d < 4$.

A.1.2 Asymptotic distribution of the empirical process

The asymptotic distribution of the empirical process Ψ_T is given in Lemma A.1 below, which is proved in Appendix B in the Supplemental Material. The proof uses consistency and asymptotic normality of kernel estimators [e.g. Bosq (1998)], the Liapunov CLT [Billingsley (1965)], results on kernel M-estimators [Tenreiro (1995)], weak convergence of empirical processes [Pollard (1990)], and a proof of stochastic equicontinuity similar to Andrews (1991).

Lemma A.1: *Under Assumptions A.1, A.4-A.19:* $\Psi_T \Longrightarrow b + \Psi$, *where* $\Psi(\theta)$, $\theta \in \Theta$, *denotes the zero-mean Gaussian stochastic process on* Θ *with covariance function given by:*

$$V_0(\theta,\tau) = E\left[\Psi(\theta)\Psi(\tau)'\right] = \begin{pmatrix} S_0(\theta,\tau) & 0 \\ 0 & w^2\Sigma_0(\theta,\tau)/f_X(x_0) \end{pmatrix}, \quad \text{for } \theta,\tau \in \Theta,$$

with:

$$S_0(\theta,\tau) = \sum_{k=-\infty}^{\infty} Cov\left[g_1(X_t,Y_t;\theta), g_1(X_{t-k},Y_{t-k};\tau)\right], \qquad \Sigma_0(\theta,\tau) = Cov\left[g_2^*(Y_t;\theta), g_2^*(Y_t;\tau)\,|X_t = x_0\right],$$

and continuous function b *is given by*

$$b(\theta) = \frac{\sqrt{\lim T h_T^{d+2m}}}{m!} w_m \frac{1}{f_X(x_0)} \begin{pmatrix} 0 \\ \Delta^m \varphi(x_0;\theta) - \frac{\varphi(x_0;\theta)}{f_X(x_0)}\Delta^m f_X(x_0) \end{pmatrix}, \quad \theta \in \Theta,$$

with $\varphi(x;\theta) := E\left[g_2^*(Y_t;\theta)|X_t = x\right] f_X(x)$, $w_m := \int_{\mathbb{R}} v^m \kappa(v)dv$. *In particular* $\Psi_T(\theta_0) \xrightarrow{d} N(\sqrt{\bar{c}}\, b_0, V_0)$ *where*

$$b_0 = \frac{w_m}{m!}\frac{1}{f_X(x_0)} \begin{pmatrix} 0 \\ \Delta^m \varphi(x_0;\theta_0) - \frac{\varphi(x_0;\theta_0)}{f_X(x_0)}\Delta^m f_X(x_0) \end{pmatrix}, \quad V_0 = \begin{pmatrix} S_0 & 0 \\ 0 & w^2\Sigma_0/f_X(x_0) \end{pmatrix}, \quad \text{(A.2)}$$

matrices S_0, Σ_0 *are defined in Assumption A.10, and* $\bar{c} := \lim T h_T^{d+2m}$.

The block diagonal elements of matrix V_0 are the standard asymptotic variance-covariance matrices of sample average and kernel regression estimators, respectively. The bias function $b(\theta)$ is zero for the unconditional moments, and is equal to the kernel regression bias for the conditional moments. Lemma A.1 implies that unconditional and conditional empirical moment restrictions are asymptotically independent, and that the convergence is uniform w.r.t. $\theta \in \Theta$.

A.1.3 Consistency of kernel moment estimators

By using the weak convergence of process $\Psi_T(\theta)$ from Lemma A.1, the Continuous Mapping Theorem, and Assumptions A.4 and A.19, we get $Q_T(\theta^*) = m_T(\theta^*)'\Omega m_T(\theta^*) + O_p(\sqrt{T})$, uniformly in $\theta^* \in \Theta \times B$. From global identification Assumption A.2, and Assumption A.20 on the weighting matrix, we have:

$$\inf_{\theta^* \in \Theta \times B: \left\|\theta^* - \theta_0^*\right\| \geq \varepsilon} m_T(\theta^*)'\Omega m_T(\theta^*) \geq C T h_T^d,$$

for any $\varepsilon > 0$ and a constant $C = C_\varepsilon > 0$. Then, by using $\sqrt{T} = o\left(T h_T^d\right)$ from Assumption A.9, we get $\inf_{\theta^* \in \Theta \times B: \left\|\theta^* - \theta_0^*\right\| \geq \varepsilon} Q_T(\theta^*) \geq \frac{1}{2} C T h_T^d$ with probability approaching 1. Since $T h_T^d \to \infty$ from Assumption A.9, we conclude that the minimizer $\widehat{\theta}_T^*$ of Q_T is such that $P\left[\left\|\widehat{\theta}_T^* - \theta_0^*\right\| \geq \varepsilon\right] \to 0$ as $T \to \infty$, for any $\varepsilon > 0$ (see Appendix B in the Supplemental Material for a detailed derivation).

A.1.4 Asymptotic distribution of kernel moment estimators

From the first-order condition $\partial Q_T\left(\widehat{\theta}_T^*\right)/\partial\theta^* = 0$ and a mean-value expansion we have:

$$\frac{\partial \widehat{g}_T'}{\partial \theta^*}\left(\widehat{\theta}_T^*\right)\Omega \widehat{g}_T\left(\theta_0^*\right) + \frac{\partial \widehat{g}_T'}{\partial \theta^*}\left(\widehat{\theta}_T^*\right)\Omega \frac{\partial \widehat{g}_T}{\partial \theta^{*\prime}}\left(\widetilde{\theta}_T^*\right)\left(\widehat{\theta}_T^* - \theta_0^*\right) = 0, \tag{A.3}$$

where $\widetilde{\theta}_T^*$ is between $\widehat{\theta}_T^*$ and θ_0^* componentwise. Compared to the standard GMM framework, we have to disentangle the linear transformations of θ_0 with parametric and nonparametric convergence rates. Let us introduce the invertible $(p+L, p+L)$ matrix:

$$R_T = \begin{pmatrix} T^{-1/2}R_{1,Z} & \left(Th_T^d\right)^{-1/2} R_{2,Z} & 0 \\ 0 & 0 & \left(Th_T^d\right)^{-1/2} Id_L \end{pmatrix},$$

where the matrices $R_{1,Z}$ and $R_{2,Z}$ are defined by the linear change of parameter from θ to $\eta = (\eta_1', \eta_2')'$ in (3.13). By pre-multiplying equation (A.3) by matrix R_T' we get:

$$R_T'\frac{\partial \widehat{g}_T'}{\partial \theta^*}\left(\widehat{\theta}_T^*\right)\Omega \widehat{g}_T\left(\theta_0^*\right) + R_T'\frac{\partial \widehat{g}_T'}{\partial \theta^*}\left(\widehat{\theta}_T^*\right)\Omega \frac{\partial \widehat{g}_T}{\partial \theta^{*\prime}}\left(\widetilde{\theta}_T^*\right) R_T \begin{pmatrix} \sqrt{T}\left(\widehat{\eta}_{1,T} - \eta_{1,0}\right) \\ \sqrt{Th_T^d}\left(\widehat{\eta}_{2,T} - \eta_{2,0}\right) \\ \sqrt{Th_T^d}\left(\widehat{\beta}_T - \beta_0\right) \end{pmatrix} = 0.$$

We have the following Lemma A.2, proved in Appendix B in the Supplemental Material using the ULLN and the CLT for mixing processes in Potscher, Prucha (1989), and Herrndorf (1984), respectively.

Lemma A.2: *Under Assumptions A.1-A.24 we have* $\text{plim}_{T\to\infty} \frac{\partial \widehat{g}_T}{\partial \theta^{*\prime}}\left(\widehat{\theta}_T^*\right) R_T = \text{plim}_{T\to\infty} \frac{\partial \widehat{g}_T}{\partial \theta^{*\prime}}\left(\widetilde{\theta}_T^*\right) R_T = J_0$, *where matrix* J_0 *is given by:*

$$J_0 = \begin{pmatrix} E\left(\frac{\partial g_1}{\partial \theta'}\right) R_{1,Z} & 0 & 0 \\ 0 & E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_{2,Z} & 0 \\ 0 & E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R_{2,Z} & -Id_L \end{pmatrix} = \begin{pmatrix} E\left(\frac{\partial g_1}{\partial \eta_1'}\right) & 0 & 0 \\ 0 & E\left(\frac{\partial g_2}{\partial \eta_2'}|x_0\right) & 0 \\ 0 & E\left(\frac{\partial a}{\partial \eta_2'}|x_0\right) & -Id_L \end{pmatrix}. \tag{A.4}$$

From the local identification Assumption A.3, matrix J_0 is full rank. Thus:

$$\left(\sqrt{T}\left(\widehat{\eta}_{1,T} - \eta_{1,0}\right)', \sqrt{Th_T^d}\left(\widehat{\eta}_{2,T} - \eta_{2,0}\right)', \sqrt{Th_T^d}\left(\widehat{\beta}_T - \beta_0\right)'\right)' = -\left(J_0'\Omega J_0\right)^{-1} J_0'\Omega \widehat{g}_T\left(\theta_0^*\right) + o_p(1). \tag{A.5}$$

Since $\widehat{g}_T\left(\theta_0^*\right) = \Psi_T(\theta_0) \xrightarrow{d} N\left(\sqrt{c}\, b_0, V_0\right)$ from Lemma A.1, Proposition 2 follows.

A.1.5 Optimal weighting matrix for given instrument and bandwidth

When the bandwidth is such that $h_T = cT^{-1/(2m+d)}$, for some constant $c > 0$, from Proposition 2 the asymptotic MSE of $\left(\sqrt{T}\left(\widehat{\eta}_{1,T} - \eta_{1,0}\right)', \sqrt{Th_T^d}\left(\widehat{\eta}_{2,T} - \eta_{2,0}\right)', \sqrt{Th_T^d}\left(\widehat{\beta}_T - \beta_0\right)'\right)'$ is $\left(J_0'\Omega J_0\right)^{-1}$ $J_0'\Omega M_0 \Omega J_0\left(J_0'\Omega J_0\right)^{-1}$, where $M_0 := V_0 + c^{2m+d} b_0 b_0'$. The optimal weighting matrix for given instrument Z and bandwidth constant c is:

$$\Omega = M_0^{-1} = \left(V_0 + c^{2m+d} b_0 b_0'\right)^{-1}. \tag{A.6}$$

The corresponding minimal MSE is $\left(J_0^{'}M_0^{-1}J_0\right)^{-1}$. Since M_0 and J_0 are block diagonal w.r.t. η_1 and $\left(\eta_2^{'},\beta^{'}\right)^{'}$, the associated asymptotic MSE of the estimator of β is:

$$M\left(Z,c,a\right)=e^{'}\left(J_{0,Z}^{'}\left(\frac{w^2}{c^d f_X(x_0)}\Sigma_0+c^{2m}w_m^2 b(x_0)b(x_0)^{'}\right)^{-1}J_{0,Z}\right)^{-1}e, \tag{A.7}$$

where $e=(\underset{L\times s_Z}{0},Id_L)^{'}$, $b\left(x\right)=\frac{1}{m!}\left(\frac{\Delta^m\varphi(x;\theta_0)}{f_X(x)}-\frac{\varphi(x;\theta_0)}{f_X(x)^2}\Delta^m f_X(x)\right)$ and

$$J_{0,Z}=:\left(\begin{array}{cc} E\left(\frac{\partial g_2}{\partial\theta^{'}}|x_0\right)R_{2,Z} & 0\\ E\left(\frac{\partial a}{\partial\theta^{'}}|x_0\right)R_{2,Z} & -Id_L\end{array}\right)=\left(\begin{array}{cc} E\left(\frac{\partial g_2}{\partial\eta_2^{'}}|x_0\right) & 0\\ E\left(\frac{\partial a}{\partial\eta_2^{'}}|x_0\right) & -Id_L\end{array}\right).$$

A.1.6 Proof of Propositions 3 and 5

The optimal instrument Z^* and bandwidth constant c^* are derived by minimizing function $M\left(Z,c,a\right)$ in (A.7) w.r.t. Z and c. In the standard GMM framework with full-information identified parameter θ_0, the matrix $R_{2,Z}$ is empty, and $M\left(Z,c,a\right)$ is independent of Z. Any admissible instrument is optimal for estimating the conditional moment β, and c^* corresponds to the usual optimal bandwidth constant for kernel regression estimation [e.g., Silverman (1986)]. In the full-information underidentified case, the optimal instrument and bandwidth selection problems are nonstandard, as seen below.

i) Optimal instruments

Let us first prove that instrument Z^* in (3.14) is admissible, that is, satisfies Assumptions A.2 and A.3. We have, for any vector $\alpha\in\mathbb{R}^p$:

$$\begin{aligned} E\left[\frac{\partial g_1}{\partial\theta^{'}}\left(X,Y;\theta_0\right)\right]\alpha &= 0\Longleftrightarrow E\left[E\left(\frac{\partial g^{'}}{\partial\theta}\left(Y;\theta_0\right)|X\right)W(X)E\left(\frac{\partial g}{\partial\theta^{'}}\left(Y;\theta_0\right)|X\right)\alpha\right]=0\\ &\Longleftrightarrow \alpha^{'}E\left(\frac{\partial g^{'}}{\partial\theta}\left(Y;\theta_0\right)|X\right)W(X)E\left(\frac{\partial g}{\partial\theta^{'}}\left(Y;\theta_0\right)|X\right)\alpha=0,\quad P_0\text{-a.s.},\\ &\Longleftrightarrow E\left(\frac{\partial g}{\partial\theta^{'}}\left(Y;\theta_0\right)|X\right)\alpha=0,\quad P_0\text{-a.s.}. \end{aligned} \tag{A.8}$$

Thus, Assumption A.3 follows from Assumption a.2 in the text. Then, Assumption A.2 is also satisfied if Θ is taken small enough, which is sufficient for the validity of the asymptotic results. From Assumption A.25, the asymptotic properties in Proposition 2 apply for the kernel moment estimators with instrument Z^*.

Let us now prove that instrument Z^* is optimal. In the expression of $M\left(Z,c,a\right)$ in (A.7), instrument Z affects matrix $J_{0,Z}$ only, and the matrix $J_{0,Z}$ depends on Z through $\mathcal{J}_Z=\text{Ker}\,E\left[\partial g_1\left(X,Y;\theta_0\right)/\partial\theta^{'}\right]$, only. If Z and $\widetilde{Z}$ are two admissible instruments such that $\mathcal{J}_Z\subset\mathcal{J}_{\widetilde{Z}}$, then $M\left(Z,c,a\right)\leq M\left(\widetilde{Z},c,a\right)$. Since $\mathcal{J}^*\subset\mathcal{J}_Z$ for any admissible instrument Z, Z^* is an optimal instrument if $\mathcal{J}^*=\mathcal{J}_{Z^*}$. The latter equality follows from (A.8).

Since $\mathcal{J}^*=\mathcal{J}_{Z^*}$, the ranges of matrix R_2 defined in (3.5) and matrix R_{2,Z^*} coincide. From (A.7), the asymptotic MSE for (any) optimal instrument Z^* becomes

$$M(c,a)=e^{'}\left(J_0^{*'}\left(\frac{1}{c^d}\frac{w^2}{f_X(x_0)}\Sigma_0+c^{2m}w_m^2 b(x_0)b(x_0)^{'}\right)^{-1}J_0^*\right)^{-1}e,$$

where $J_0^* = \left(\begin{array}{cc} E\left(\frac{\partial g_2}{\partial \eta_2^{*\prime}}|x_0\right) & 0 \\ E\left(\frac{\partial a}{\partial \eta_2^{*\prime}}|x_0\right) & -Id_L \end{array} \right)$.

ii) Optimal bandwidth

In the rest of this proof we assume $L = \dim(a) = 1$. Then, function $M(c, a)$ is scalar, and the first-order condition for minimizing $M(c, a)$ w.r.t. c is given by:

$$\frac{\partial M(c,a)}{\partial c} = e^{'} \left[\left(J_0^{*'} \tilde{\Sigma}^{-1} J_0^* \right)^{-1} J_0^{*'} \tilde{\Sigma}^{-1} \left(-\frac{d}{c^{d+1}} \frac{w^2}{f_X(x_0)} \Sigma_0 + 2mc^{2m-1} w_m^2 b(x_0) b(x_0)^{'} \right) \tilde{\Sigma}^{-1} J_0^* \left(J_0^{*'} \tilde{\Sigma}^{-1} J_0^* \right)^{-1} \right] e = 0$$

with $\tilde{\Sigma} := \frac{1}{c^d} \frac{w^2}{f_X(x_0)} \Sigma_0 + c^{2m} w_m^2 b(x_0)^{'} b(x_0)$ and $e = (0^{'}, 1)^{'} \in \mathbb{R}^{s^*} \times \mathbb{R}$. The solution $c^*(a)$ is $c^*(a) = \xi^{1/(2m+d)}$ where $\xi = \xi(a)$ satisfies the equation:

$$\xi \quad = \quad \frac{w^2 d}{2mw_m^2 f_X(x_0)} \frac{e^{'} \left[\left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} J_0^{*'} A(\xi)^{-1} \Sigma_0 A(\xi)^{-1} J_0^* \left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} \right] e}{e^{'} \left[\left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} J_0^{*'} A(\xi)^{-1} b(x_0) b(x_0)^{'} A(\xi)^{-1} J_0^* \left(J_0^{*'} A(\xi)^{-1} J_0^* \right)^{-1} \right] e}, \tag{A.9}$$

where $A(\xi) := \frac{w^2}{f_X(x_0)} \Sigma_0 + \xi b(x_0) b(x_0)^{'}$. The optimal bandwidth sequence is $h_T = c^*(a) T^{-1/(2m+d)}$. Then, from (A.6) the optimal weighting matrix becomes:

$$\Omega^*(a) = \left(V_0 + c^{*2m+d} b_0 b_0^{'} \right)^{-1}, \tag{A.10}$$

where V_0 is defined in (A.2) with optimal instrument Z^*. This concludes the proof of Proposition 3.

iii) Kernel nonparametric efficiency bound

The kernel nonparametric efficiency bound is $M(c^*(a), a)$ and is equal to $\mathcal{M}(x_0, a)$ in Proposition 5.

A.1.7 Proof of Corollary 6

The proof of Corollary 6 is similar to the proof of Propositions 3 and 5, and is given in Appendix B in the Supplemental Material. In particular, the optimal weighting matrix is $\Omega = V_0^{-1}$.

A.2 Asymptotic properties of the XMM estimator

A.2.1 Concentration with respect to the functional parameters

We first concentrate the estimation criterion in Definition 1 w.r.t. the functional parameters. Let us introduce the Lagrange multipliers $\lambda, \mu, \lambda_t, \mu_t$, $t = 1, ..., T$. The Lagrangian function is given by:

$$\begin{aligned} \mathcal{L}_T \quad = \quad & \frac{1}{T} \sum_{t=1}^T \int \frac{\left[\widehat{f}(y|x_t) - f_t(y) \right]^2}{\widehat{f}(y|x_t)} dy + h_T^d \int \log \left[f_0(y) / \widehat{f}(y|x_0) \right] f_0(y) dy \\ & -2 \frac{1}{T} \sum_{t=1}^T \mu_t \left(\int f_t(y) dy - 1 \right) - h_T^d \mu \left(\int f_0(y) dy - 1 \right) \\ & -2 \frac{1}{T} \sum_{t=1}^T \lambda_t^{'} \int g(y; \theta) f_t(y) dy - h_T^d \lambda^{'} \int g_2(y; \theta) f_0(y) dy. \end{aligned} \tag{A.11}$$

The first-order conditions w.r.t. the functional parameters f_t, $t = 1, ..., T$, and f_0 yield:

$$
\begin{aligned}
f_t(y) &= \widehat{f}(y|x_t) + \mu_t \widehat{f}(y|x_t) + \lambda_t^{'} g(y;\theta)\widehat{f}(y|x_t), \quad t = 1, ..., T, & \text{(A.12)}\\
f_0(y) &= \widehat{f}(y|x_0) \exp\left(\lambda^{'} g_2(y;\theta) + \mu - 1\right). & \text{(A.13)}
\end{aligned}
$$

The Lagrange multipliers λ_t, μ_t, $t = 1, ..., T$, and μ can be deduced from the constraints. We get:

$$
\lambda_t = -\widehat{V}\left(g(\theta)|x_t\right)^{-1} \widehat{E}\left(g(\theta)|x_t\right), \ \ \mu_t = -\lambda_t^{'} \widehat{E}\left(g(\theta)|x_t\right), \ \ exp(1-\mu) = \widehat{E}\left[\exp\left(\lambda^{'} g_2(\theta)\right)|x_0\right], \tag{A.14}
$$

where $\widehat{E}(.|x)$ and $\widehat{V}(.|x)$ denote the conditional expectation and the conditional variance w.r.t. the kernel density estimator $\widehat{f}(.|x_t)$, respectively. By replacing (A.14) into (A.12)-(A.13) we get the concentrated functional parameters:

$$
\begin{aligned}
f_t(y;\theta) &= \widehat{f}(y|x_t) - \widehat{E}\left(g(\theta)|x_t\right)^{'} \widehat{V}\left(g(\theta)|x_t\right)^{-1} \left[g(y;\theta) - \widehat{E}\left(g(\theta)|x_t\right)\right] \widehat{f}(y|x_t), \ t = 1, ..., T,\\
f_0(y;\theta,\lambda) &= \frac{\exp\left(\lambda^{'} g_2(y;\theta)\right)}{\widehat{E}\left[\exp\left(\lambda^{'} g_2(\theta)\right)|x_0\right]} \widehat{f}(y|x_0). & \text{(A.15)}
\end{aligned}
$$

The concentrated Lagrangian becomes:

$$
\mathcal{L}_T^c(\theta,\lambda) = \frac{1}{T}\sum_{t=1}^{T} \widehat{E}\left(g(\theta)|x_t\right)^{'} \widehat{V}\left(g(\theta)|x_t\right)^{-1} \widehat{E}\left(g(\theta)|x_t\right) - h_T^d \log \widehat{E}\left(\exp\left(\lambda^{'} g_2(\theta)\right)|x_0\right). \tag{A.16}
$$

The XMM estimator $\widehat{\theta}_T$ is the minimizer of $\mathcal{L}_T^c(\theta) = \mathcal{L}_T^c(\theta, \lambda(\theta))$, where $\lambda(\theta) = \arg\max_\lambda \mathcal{L}_T^c(\theta,\lambda)$ is as such that $\widehat{E}\left[g_2(\theta)\exp\left(\lambda(\theta)^{'} g_2(\theta)\right)|x_0\right] = 0$ for any θ [see (2.11)-(2.12)]. The estimator $\widehat{f}^*(.|x_0)$ of $f(.|x_0)$ is obtained from (A.15) by replacing θ by $\widehat{\theta}_T$, and λ by $\widehat{\lambda}_T = \lambda\left(\widehat{\theta}_T\right)$ [see (2.13)].

A.2.2 Asymptotic expansions

i) Asymptotic expansion of estimator $\widehat{\theta}_T$ and Lagrange multiplier $\widehat{\lambda}_T$

The asymptotic expansion of estimator $\widehat{\theta}_T$ and Lagrange multiplier $\widehat{\lambda}_T$ is derived by expanding the criterion (A.16) around $\theta = \theta_0$ and $\lambda = 0$. We have to distinguish between the linear combinations of $\widehat{\theta}_T$ converging at a parametric rate and those converging at a nonparametric rate. For this purpose, we use the change of parameterization in (3.5) from θ to $\eta^* = (\eta_1^{*\prime}, \eta_2^{*\prime})^{\prime}$. The asymptotic expansion for the estimators of the transformed parameters $\widehat{\eta}_{1,T}^*$, $\widehat{\eta}_{2,T}^*$ and the Lagrange multiplier $\widehat{\lambda}_T$ are given in Lemma A.3 below.

Lemma A.3: *(i) The asymptotic expansions of $\widehat{\eta}_{1,T}^*$ and $\widehat{\eta}_{2,T}^*$ are given by:*

$$
\begin{aligned}
\sqrt{T}\left(\widehat{\eta}_{1,T}^* - \eta_{1,0}^*\right) &\simeq -\left(R_1^{\prime} E\left[E\left(\frac{\partial g^{'}}{\partial\theta}|X\right) V\left(g|X\right)^{-1} E\left(\frac{\partial g}{\partial\theta^{'}}|X\right)\right] R_1\right)^{-1}\\
&\quad \cdot \frac{1}{\sqrt{T}}\sum_{t=1}^{T} R_1^{\prime} E\left(\frac{\partial g^{'}}{\partial\theta}|x_t\right) V\left(g|x_t\right)^{-1} g(y_t;\theta_0),
\end{aligned}
$$

and:

$$
\begin{aligned}
\sqrt{Th_T^d}\left(\widehat{\eta}_{2,T}^* - \eta_{2,0}^*\right) &\simeq -\left[R_2^{\prime} E\left(\frac{\partial g_2^{'}}{\partial\theta}|x_0\right) V\left(g_2|x_0\right)^{-1} E\left(\frac{\partial g_2}{\partial\theta^{'}}|x_0\right) R_2\right]^{-1}\\
&\quad \cdot R_2^{\prime} E\left(\frac{\partial g_2^{'}}{\partial\theta}|x_0\right) V\left(g_2|x_0\right)^{-1} \sqrt{Th_T^d}\int g_2(y;\theta_0)\widehat{f}(y|x_0)dy, & \text{(A.17)}
\end{aligned}
$$

respectively. (ii) The asymptotic expansion of $\widehat{\lambda}_T$ is:

$$\widehat{\lambda}_T \simeq -V\left(g_2|x_0\right)^{-1}(Id-M)\int g_2(y;\theta_0)\widehat{f}(y|x_0)dy, \tag{A.18}$$

where M is the orthogonal projection matrix on the column space of $E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right)R_2$ for the inner product defined by $V\left(g_2|x_0\right)^{-1}$:

$$M = E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right)R_2\left[R_2'E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right)V\left(g_2|x_0\right)^{-1}E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right)R_2\right]^{-1}R_2'E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right)V\left(g_2|x_0\right)^{-1}. \tag{A.19}$$

The asymptotic expansion of $\widehat{\eta}^*_{1,T}$ comes from the part of the criterion (A.16) involving the uniform moment restrictions, since the part involving the local moment restrictions is asymptotically negligible for $\widehat{\eta}^*_{1,T}$. The asymptotic expansion of $\widehat{\eta}^*_{2,T}$ involves the local moment restrictions only, since the uniform moment restrictions are not informative for parameter η_2. The term $\log\widehat{E}\left(\exp\left(\lambda' g_2(\theta)\right)|x_0\right)$ in (A.16) is induced by the KLIC component and has a weighting factor h_T^d (see Definition 1). This weighting factor ensures that the contribution of the discrepancy measure associated with the local restrictions at x_0 is asymptotically the same as for a kernel moment estimator with optimal weighting matrix and instruments.

ii) Asymptotic expansion of $\widehat{f}^*(.|x_0)$

Using $\widehat{f}^*(y|x_0) = f_0(y;\widehat{\theta}_T,\widehat{\lambda}_T)$, from (A.15) we have:

$$\begin{aligned}\widehat{f}^*(y|x_0) &\simeq \frac{1+\widehat{\lambda}_T' g_2\left(y;\widehat{\theta}_T\right)}{1+\widehat{\lambda}_T'\widehat{E}\left(g_2(\widehat{\theta}_T)|x_0\right)}\widehat{f}(y|x_0) \simeq \left[1+\widehat{\lambda}_T'\left(g_2\left(y;\widehat{\theta}_T\right)-\widehat{E}\left(g_2(\widehat{\theta}_T)|x_0\right)\right)\right]\widehat{f}(y|x_0)\\ &\simeq \widehat{f}(y|x_0)+\widehat{\lambda}_T' g_2(y;\theta_0)f(y|x_0).\end{aligned}$$

Then, from (A.18) we get:

$$\widehat{f}^*(y|x_0) \simeq \widehat{f}(y|x_0) - f(y|x_0)g_2(y;\theta_0)'V\left(g_2|x_0\right)^{-1}(Id-M)\int g_2(y;\theta_0)\widehat{f}(y|x_0)dy. \tag{A.20}$$

iii) Asymptotic expansion of $\widehat{E}^*(a|x_0)$

We have:

$$\begin{aligned}\widehat{E}^*(a|x_0) &= \int a(y;\widehat{\theta}_T)\widehat{f}^*(y|x_0)dy\\ &\simeq \int a(y;\theta_0)f(y|x_0)dy + \int\frac{\partial a}{\partial \theta'}(y;\theta_0)f(y|x_0)dy\left(\widehat{\theta}_T-\theta_0\right)+\int a(y;\theta_0)\left[\widehat{f}^*(y|x_0)-f(y|x_0)\right]dy\\ &\simeq E(a|x_0)+E\left(\frac{\partial a}{\partial \theta'}|x_0\right)R_2\left(\widehat{\eta}^*_{2,T}-\eta^*_{2,0}\right)\\ &+\int a(y;\theta_0)\left\{\widehat{f}(y|x_0)-f(y|x_0)-f(y|x_0)g_2(y;\theta_0)'V\left(g_2|x_0\right)^{-1}(Id-M)\int g_2(y;\theta_0)\widehat{f}(y|x_0)dy\right\}dy,\end{aligned}$$

where the last asymptotic equivalence comes from (A.20) and the fact that the contribution of $\widehat{\eta}_{1,T}^* - \eta_{1,0}^*$ is asymptotically negligible. Then, from (A.17) we get:

$$\begin{aligned}\widehat{E}^*(a|x_0) \quad &\simeq \quad E(a|x_0) - E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R_2 \left[R_2' E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V\left(g_2|x_0\right)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2\right]^{-1} \\ &\cdot R_2' E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V\left(g_2|x_0\right)^{-1} \int g_2(y;\theta_0)\widehat{f}(y|x_0)dy \\ &+ \int a(y;\theta_0)\left[\widehat{f}(y|x_0) - f(y|x_0)\right] dy - Cov\left(a, g_2|x_0\right) V\left(g_2|x_0\right)^{-1} (Id - M) \int g_2(y;\theta_0)\widehat{f}(y|x_0)dy.\end{aligned}$$

We deduce the asymptotic expansion:

$$\begin{aligned}&\widehat{E}^*(a|x_0) - E(a|x_0) \simeq \int a(y;\theta_0)\delta\widehat{f}(y|x_0)dy - Cov\left(a, g_2|x_0\right) V\left(g_2|x_0\right)^{-1} \int g_2(y;\theta_0)\delta\widehat{f}(y|x_0)dy \\ &- \left[E\left(\frac{\partial a}{\partial \theta'}|x_0\right) R_2 - Cov\left(a, g_2|x_0\right) V\left(g_2|x_0\right)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2\right] \\ &\cdot \left[R_2' E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V\left(g_2|x_0\right)^{-1} E\left(\frac{\partial g_2}{\partial \theta'}|x_0\right) R_2\right]^{-1} R_2' E\left(\frac{\partial g_2'}{\partial \theta}|x_0\right) V\left(g_2|x_0\right)^{-1} \int g_2(y;\theta_0)\delta\widehat{f}(y|x_0)dy,\end{aligned} \tag{A.21}$$

where $\delta\widehat{f}(y|x_0) := \widehat{f}(y|x_0) - f(y|x_0)$.

A.2.3 Asymptotic distribution of the XMM estimator (Proofs of Proposition 7 and Corollary 8)

Let us derive the asymptotic distribution of the estimator $\widehat{E}^*(a|x_0)$. In the asymptotic expansion (A.21), the first two terms in the RHS correspond to the residual of the regression of $\int a(y;\theta_0)\delta\widehat{f}(y|x_0)dy$ on $\int g_2(y;\theta_0)\delta\widehat{f}(y|x_0)dy$. This residual is asymptotically independent of the third term in the RHS. Thus, from the asymptotic normality of integrals of kernel estimators, we get $\frac{\sqrt{Th_T^d}}{w}\left[\widehat{E}^*(a|x_0) - E(a|x_0)\right] \xrightarrow{d} N(0, \mathcal{B}(x_0,a))$, where $\mathcal{B}(x_0,a)$ is given in Corollary 6. This proves Proposition 7. Corollary 8 is proved by checking that the asymptotic expansion for the XMM estimator $\hat{\theta}_T$ in Lemma A.3 (i) corresponds to the asymptotic expansion for the kernel moment estimator in (A.5) with $\bar{c} = \lim Th_T^{d+2m} = 0$, $\Omega = V_0^{-1}$ and $Z = E\left[\frac{\partial g'}{\partial \theta}(Y;\theta_0)|X\right] V\left[g(Y;\theta_0)|X\right]^{-1}$. The details of the derivation are given in Appendix B in the Supplemental Material.

A.3 A parametric stochastic volatility model

In this Appendix we consider the parametric stochastic volatility model introduced in Section 3.2. We first derive equations (3.9) from the no-arbitrage conditions, then we give the dynamics under the risk-neutral distribution, and finally we prove Proposition 1.

A.3.1 Proof of equations (3.9)

Let us assume that the DGP P_0 is compatible with the historical dynamic of the state variables $X_t = (\tilde{r}_t, \sigma_t^2)'$ given in (3.6) and the sdf (3.7) with true parameter value θ_0. The restrictions implied by the no-arbitrage assumption for the riskfree asset and the underlying asset are given by:

$$E_0\left[M_{t,t+1}(\theta_0)\exp r_{f,t+1}|x_t\right] = 1, \quad E_0\left[M_{t,t+1}(\theta_0)\exp r_{t+1}|x_t\right] = 1, \quad \forall x_t = \left(\tilde{r}_t, \sigma_t^2\right)', \tag{A.22}$$

respectively. Let us first consider the no-arbitrage restriction for the riskfree asset. We have:

$$
\begin{aligned}
E_0\left[M_{t,t+1}(\theta_0)e^{r_{f,t+1}}|x_t\right] &= E_0\left[\exp\left(-\theta_1^0-\theta_2^0\sigma_{t+1}^2-\theta_3^0\sigma_t^2-\theta_4^0\tilde{r}_{t+1}\right)|x_t\right] \\
&= \exp\left(-\theta_1^0-\theta_3^0\sigma_t^2\right)E_0\left[\exp\left(-\left[\theta_2^0+\theta_4^0\gamma_0-\frac{(\theta_4^0)^2}{2}\right]\sigma_{t+1}^2\right)|x_t\right] \\
&= \exp\left(-\left[\theta_1^0+b_0\left(\theta_2^0+\theta_4^0\gamma_0-\frac{(\theta_4^0)^2}{2}\right)\right]-\left[\theta_3^0+a_0\left(\theta_2^0+\theta_4^0\gamma_0-\frac{(\theta_4^0)^2}{2}\right)\right]\sigma_t^2\right),
\end{aligned}
$$

where we have integrated $\tilde{r}_{t+1}$ conditional on σ_{t+1}^2 and used the definition of the ARG process. Since the RHS has to be equal to 1 for any admissible value of σ_t^2, we get:

$$
\theta_1^0=-b_0\left(\theta_2^0+\theta_4^0\gamma_0-\frac{(\theta_4^0)^2}{2}\right),\ \theta_3^0=-a_0\left(\theta_2^0+\theta_4^0\gamma_0-\frac{(\theta_4^0)^2}{2}\right). \tag{A.23}
$$

Let us now consider the no-arbitrage restriction for the underlying asset. By using $E_0\left[M_{t,t+1}(\theta_0)\exp r_{t+1}|x_t\right]=E_0\left[\exp\left(-\theta_1^0-\theta_2^0\sigma_{t+1}^2-\theta_3^0\sigma_t^2-\left(\theta_4^0-1\right)\tilde{r}_{t+1}\right)|x_t\right]$, we get the same conditions as in (A.23) by replacing θ_4^0 with θ_4^0-1:

$$
\theta_1^0=-b_0\left(\theta_2^0+(\theta_4^0-1)\gamma_0-\frac{(\theta_4^0-1)^2}{2}\right),\ \theta_3^0=-a_0\left(\theta_2^0+(\theta_4^0-1)\gamma_0-\frac{(\theta_4^0-1)^2}{2}\right).
$$

Since functions a_0 and b_0 are one-to-one, we get $\theta_2^0+\left(\theta_4^0-1\right)\gamma_0-\frac{\left(\theta_4^0-1\right)^2}{2}=\theta_2^0+\theta_4^0\gamma_0-\frac{(\theta_4^0)^2}{2}$, that is, $\theta_4^0=\gamma_0+\frac{1}{2}$. Equations (3.9) follow.

A.3.2 Risk-neutral distribution

The risk-neutral distribution Q is characterized by the conditional Laplace transform of $(\tilde{r}_{t+1},\sigma_{t+1}^2)'$, which is given by $E_0^Q\left[\exp(-u\tilde{r}_{t+1}-v\sigma_{t+1}^2)|x_t\right]=e^{-r_{f,t+1}}E_0\left[M_{t,t+1}(\theta_0)\exp(-u\tilde{r}_{t+1}-v\sigma_{t+1}^2)|x_t\right]$. In Appendix B in the Supplemental Material we show that, under the risk-neutral distribution Q, the underlying asset return follows an ARG stochastic volatility model with adjusted risk premium parameter $\gamma_0^*=-1/2$ and volatility parameters:

$$
\rho_0^*=\frac{\rho_0}{\left[1+c_0\left(\theta_2^0+\gamma_0^2/2-1/8\right)\right]^2},\quad \delta_0^*=\delta_0,\quad c_0^*=\frac{c_0}{1+c_0\left(\theta_2^0+\gamma_0^2/2-1/8\right)}. \tag{A.24}
$$

A.3.3 Proof of Proposition 1

Let us consider XMM estimation with parametric sdf $M_{t,t+1}(\theta)=e^{-r_{f,t+1}}e^{-\theta_1-\theta_2\sigma_{t+1}^2-\theta_3\sigma_t^2-\theta_4\tilde{r}_{t+1}}$. This is a well-specified sdf, with true parameter value θ_0 satisfying the restrictions in equations (3.9). The econometrician uses the uniform and local conditional moment restrictions (2.2) and (2.3) to identify θ_0. We first derive subspace $\mathcal{J}^*\subset\mathbb{R}^4$ and matrix R_2

i) Derivation of subspace $\mathcal{J}^*$ and matrix R_2

The null space $\mathcal{J}^*$ associated with the uniform conditional moment restrictions is the linear space of vectors $\alpha\in\mathbb{R}^4$ such that:

$$
E_0\left[\begin{pmatrix}\exp r_{f,t+1}\\ \exp r_{t+1}\end{pmatrix}\frac{\partial M_{t,t+1}}{\partial\theta'}(\theta_0)\mid x_t\right]\alpha=0,\quad\forall x_t. \tag{A.25}
$$

By using $\partial M_{t,t+1}(\theta_0)/\partial\theta'=M_{t,t+1}(\theta_0)\xi_{t+1}'$, we get the characterization of $\mathcal{J}^*$ given in Proposition 1 (i). Moreover, since θ_0 satisfies the no-arbitrage restrictions (A.22), we deduce that any vector

$\theta = \theta_0 + \alpha\varepsilon$, where ε is small and α satisfies (A.25), is such that $E_0\left[M_{t,t+1}(\theta) \exp r_{f,t+1}|x_t\right] = 1$ and $E_0\left[M_{t,t+1}(\theta) \exp r_{t+1}|x_t\right] = 1$, $\forall x_t$, at first-order in ε. Therefore, the vectors in $\mathcal{J}^*$ are the directions $d\theta = \theta - \theta_0$ of infinitesimal parameter changes that are compatible with no-arbitrage. From Section A.3.1, the parameter vectors θ compatible with no-arbitrage are characterized by the nonlinear restrictions $\theta_1 = -b_0(\lambda_2)$, $\theta_3 = -a_0(\lambda_2)$, $\theta_4 = \gamma_0 + 1/2$, where $\lambda_2 := \theta_2 + \gamma_0^2/2 - 1/8$. Thus, the tangent set at θ_0 is spanned by the vector:

$$\alpha = \left(d\theta_1/d\theta_2, d\theta_2/d\theta_2, d\theta_3/d\theta_2, d\theta_4/d\theta_2\right)' |_{\theta=\theta_0} = \left(-\frac{db_0}{du}(\lambda_2^0), 1, -\frac{da_0}{du}(\lambda_2^0), 0\right)' = r_2', \qquad \text{(A.26)}$$

where $\lambda_2^0 := \theta_2^0 + \gamma_0^2/2 - 1/8$. We deduce that the linear space $\mathcal{J}^*$ has dimension $\dim(\mathcal{J}^*) = 1$ and is spanned by the vector r_2. The orthogonal matrix R_2 is given by $R_2 = r_2/\left\|r_2\right\|$.

ii) Derivation of vector $E_0\left[\partial\tilde{g}(Y;\theta_0)/\partial\theta'|X = x_0\right] r_2$ and proof of local identification

The j-th element of the $(n, 1)$ vector $E_0\left[\partial\tilde{g}(Y;\theta_0)/\partial\theta'|X = x_0\right] r_2$ is given by:

$$E_0\left[\frac{\partial\tilde{g}_j(Y;\theta_0)}{\partial\theta'}|X = x_0\right] r_2 = E_0\left[\frac{\partial M_{t,t+h_j}(\theta_0)}{\partial\theta'}\left(e^{R_{t,h_j}} - k_j\right)^+ |X_t = x_0\right] r_2, \quad j = 1, \cdots, n.$$

By using $\partial M_{t,t+h}(\theta_0)/\partial\theta' = M_{t,t+h}(\theta_0)\sum_{l=1}^{h}\xi_{t+l}'$, $r_2 = (-\delta_0^* c_0^*, 1, -\rho_0^*, 0)'$ from (A.24), and:

$$\xi_{t+1}' r_2 = \sigma_{t+1}^2 - \left(\rho_0^*\sigma_t^2 + \delta_0^* c_0^*\right) = \sigma_{t+1}^2 - E_0^Q\left[\sigma_{t+1}^2|\sigma_t^2\right], \qquad \text{(A.27)}$$

we get:

$$E_0\left[\frac{\partial\tilde{g}_j(Y;\theta_0)}{\partial\theta'}|X = x_0\right] r_2 = \sum_{l=1}^{h_j} B(t_0, t_0+h_j) E_0^Q\left[\left(\sigma_{t+l}^2 - E_0^Q[\sigma_{t+l}^2|\sigma_{t+l-1}^2]\right)\left(e^{R_{t,h_j}} - k_j\right)^+ |X_t = x_0\right].$$

By conditioning on the volatility path and using the Hull-White formula [Hull, White (1987)]:

$$\begin{aligned} &B(t_0, t_0+h) E_0^Q\left[\left(\sigma_{t+l}^2 - E_0^Q[\sigma_{t+l}^2|\sigma_{t+l-1}^2]\right)\left(e^{R_{t,h}} - k\right)^+ |X_t = x_0\right] \\ &= E_0^Q\left[\left(\sigma_{t+l}^2 - E_0^Q[\sigma_{t+l}^2|\sigma_{t+l-1}^2]\right) BS(k, \sigma_{t,t+h}^2)|X_t = x_0\right], \end{aligned}$$

where $BS(k, \sigma^2)$ is the Black-Scholes price for time-to-maturity 1 and $\sigma_{t,t+h}^2$ is the integrated volatility between t and $t+h$. Moreover, since $E_0^Q\left[\sigma_{t+l}^2 - E_0^Q[\sigma_{t+l}^2|\sigma_{t+l-1}^2]|X_t\right] = E_0^Q\left[\sigma_{t+l}^2 - E_0^Q[\sigma_{t+l}^2|\sigma_{t+l-1}^2]|X_{t+l-1}\right] = 0$ by iterated expectation and the Markov property under Q, we have:

$$\begin{aligned} &E_0^Q\left[\left(\sigma_{t+l}^2 - E_0^Q[\sigma_{t+l}^2|\sigma_{t+l-1}^2]\right) BS(k, \sigma_{t,t+h}^2)|X_t = x_0\right] \\ &= Cov_0^Q\left[\sigma_{t+l}^2 - E_0^Q[\sigma_{t+l}^2|\sigma_{t+l-1}^2], BS(k, \sigma_{t,t+h}^2)|X_t = x_0\right] = Cov_0^Q\left[\sigma_{t+l}^2 - \rho_0^*\sigma_{t+l-1}^2, BS(k, \sigma_{t,t+h}^2)|X_t = x_0\right]. \end{aligned}$$

Thus, by using $\sum_{l=1}^{h_j}\left(\sigma_{t+l}^2 - \rho_0^*\sigma_{t+l-1}^2\right) = (1-\rho_0^*)\sigma_{t,t+h_j}^2 - \rho_0^*\sigma_t^2 + \rho_0^*\sigma_{t+h_j}^2$, we get:

$$\begin{aligned} &E_0\left[\partial\tilde{g}_j(Y;\theta_0)/\partial\theta'|X = x_0\right] r_2 \\ &= (1-\rho_0^*) Cov_0^Q\left(\sigma_{t,t+h_j}^2, BS(k_j, \sigma_{t,t+h_j}^2)|X_t = x_0\right) + \rho_0^* Cov_0^Q\left(\sigma_{t+h_j}^2, BS(k_j, \sigma_{t,t+h_j}^2)|X_t = x_0\right). \end{aligned}$$

Finally, let us prove that $E_0\left[\partial\tilde{g}_j(Y;\theta_0)/\partial\theta'|X = x_0\right] R_2 > 0$, for $j = 1, \cdots, n$, which implies the local identification of θ_0 (Assumption a.2). We have $Cov_0^Q\left(\sigma_{t,t+h}^2, BS(k, \sigma_{t,t+h}^2)|X_t = x_0\right) > 0$, for any $h, k > 0$, since the Black-Scholes price is strictly increasing w.r.t. the volatility and the risk-neutral distribution of $\sigma_{t,t+h}^2$ given $X_t = x_0$ is non-degenerate. Moreover we have the following Lemma A.4.

Lemma A.4: *The integrated volatility* $\sigma^2_{t,t+h}$ *is stochastically increasing in the spot volatility* σ^2_{t+h} *under the conditional risk-neutral distribution* Q *given* $X_t = x_0$*, that is,* $P_0^Q\left[\sigma^2_{t,t+h} \geq z | \sigma^2_{t+h} = s, X_t = x_0\right]$ *is increasing w.r.t.* s*, for any* z*.*

Since $BS(k, \sigma^2)$ is an increasing function of σ^2, Lemma A.4 implies that $E_0^Q\left[BS(k, \sigma^2_{t,t+h}) | \sigma^2_{t+h}, X_t = x_0\right]$ is an increasing function of σ^2_{t+h}. Thus:

$$Cov_0^Q\left(\sigma^2_{t+h}, BS(k, \sigma^2_{t,t+h}) | X_t = x_0\right) = Cov_0^Q\left(\sigma^2_{t+h}, E_0^Q\left[BS(k, \sigma^2_{t,t+h}) | \sigma^2_{t+h}, X_t = x_0\right] | X_t = x_0\right) \geq 0,$$

for any $h, k > 0$. The conclusion follows.